

大数据与人工智能

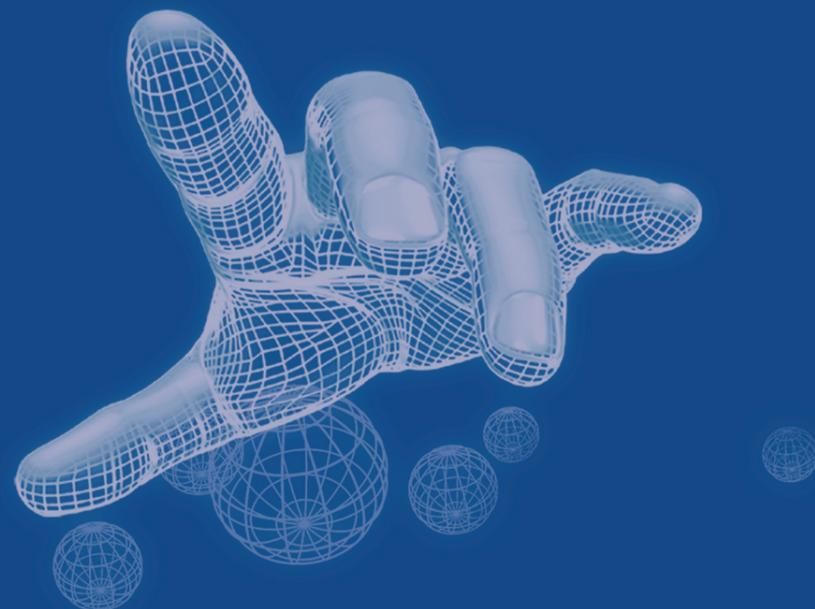
Big Data and Artificial Intelligence

Volume 5 • Issue 5 • May 2024 • ISSN 2737-4726(Online) 2737-4718(Print)

大数据与人工智能

Big Data and Artificial Intelligence

Volume 5 • Issue 5 • May 2024 • ISSN 2737-4726(Online) 2737-4718(Print)





本刊旨在汇聚国际领域的专家学者、科研人员及行业精英，共同探索大数据处理、分析、挖掘以及人工智能算法、模型、应用等方面的最新进展，分享跨学科的研究成果与实战经验，促进大数据与人工智能技术的深度融合与协同发展。

为满足广大科研人员的需要，《大数据与人工智能》期刊文章收录范围包括但不限于：

- 数据挖掘分析
- 人脸识别
- 数据工程
- 数据处理
- VR/AR
- 人工智能应用
- 数据库
- 信息统计与分析
- 数据结构
- 人工智能算法原理

SYNERGY PUBLISHING PTE. LTD.

12 Eu Tong Sen Street

#07-169

Singapore 059819



版权声明/Copyright

协同出版社出版的电子版和纸质版等文章和其他辅助材料，除另作说明外，作者有权依据Creative Commons国际署名—非商业使用4.0版权对于引用、评价及其他方面的要求，对文章进行公开使用、改编和处理。读者在分享及采用本刊文章时，必须注明原文作者及出处，并标注对本刊文章所进行的修改。关于本刊文章版权的最终解释权归协同出版社所有。

All articles and any accompanying materials published by Synergy Publishing on any media (e.g. online, print etc.), unless otherwise indicated, are licensed by the respective author(s) for public use, adaptation and distribution but subjected to appropriate citation, crediting of the original source and other requirements in accordance with the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) license. In terms of sharing and using the article(s) of this journal, user(s) must mark the author(s) information and attribution, as well as modification of the article(s). Synergy Publishing Pte. Ltd. reserves the final interpretation of the copyright of the article(s) in this journal.

About the Publisher

Synergy Publishing Pte. Ltd. (SP) is an international publisher of online, open access and scholarly peer-reviewed journals covering a wide range of academic disciplines including science, technology, medicine, engineering, education and social science. Reflecting the latest research from a broad sweep of subjects, our content is accessible worldwide – both in print and online.

SP aims to provide an analytics as well as platform for information exchange and discussion that help organizations and professionals in advancing society for the betterment of mankind. SP hopes to be indexed by well-known databases in order to expand its reach to the science community, and eventually grow to be a reputable publisher recognized by scholars and researchers around the world.

SP adopts the Open Journal Systems, see on <http://ojs.s-p.sg>

Database Inclusion



Asia & Pacific Science
Citation Index



Creative Commons



MyScienceWork



Google Scholar



Crossref



China National Knowledge
Infrastructure

大数据与人工智能

Big Data and Artificial Intelligence

Volume 5 Issue 5 May 2024
ISSN 2737-4726 (Online) 2737-4718 (Print)

主 编

陈学松
广东工业大学，中国

编 委

孙玉春 Yuchun Sun
洪一玮 Yiwei Hong
陈睿豪 Ruihao Chen
付吉颢 Jihao Fu
路林艳 Linyan Lu

1	新时代下人工智能在网络安全运维服务中的相关运用 / 袁彬	1	The Application of Artificial Intelligence in Network Security Operation and Maintenance Services in the New Era / Bin Yuan
4	面向云计算的大数据存储与计算资源动态调度算法研究 / 霍晨宇	4	Research on the Dynamic Scheduling Algorithm of Big Data Storage and Computing Resources for Cloud Computing / Chenyu Huo
7	基于大数据技术的分布式文件系统设计与实现 / 刘艳茹	7	Design and Implementation of a Distributed File System Based on Big Data Technology / Yanru Liu
10	基于嵌入式设计的软硬件开发策略思考 / 陈德娟	10	Reflection on Software and Hardware Development Strategy Based on Embedded Design / Dejuan Chen
13	网络数据挖掘在商业决策中的应用现状及改善策略 / 王帅	13	Application Status and Improvement Strategy of Network Data Mining in Business Decision-making / Shuai Wang
16	基于智能化运维的 IDC 数据中心管理探究 / 王伟奇	16	Research on IDC Data Center Management Based on Intelligent Operation and Maintenance / Weiqi Wang
19	基于大数据技术网络舆情分析系统 / 王丽颖	19	Network Public Opinion Analysis System Based on Big Data Technology / Liying Wang
22	基于人工智能的网络故障预测与诊断技术研究 / 桑晨	22	Research on Network Fault Prediction and Diagnosis Technology Based on Artificial Intelligence / Chen Sang
25	基于大数据时代视角下人工智能实践 / 张纯	25	AI Practice from the Perspective of the Era of Big Data / Chun Zhang
28	云计算环境中大数据的实时安全监控与审计技术研究 / 孔淼	28	Research on Real-time Security Monitoring and Auditing Technology of Big Data in Cloud Computing Environment / Miao Kong
31	高等教育自学考试实践考核中开放式编程题的抄袭检测方法研究 / 李睿 张轶雄 郭强 王昱 白雁	31	Research on Plagiarism Detection Methods for Open-ended Programming Questions in Higher Education Self-study Exams / Rui Li Yixiong Zhang Qiang Guo Yu Wang Yan Bai
37	人工智能在智能制造中的应用研究 / 马广建	37	Research on the Application of Artificial Intelligence in Intelligent Manufacturing / Guangjian Ma
40	数字孪生技术在军事领域的应用分析 / 范雪娇 金双	40	Analysis of Digital Twin Technology Applications in the Military Field / Xuejiao Fan Shuang Jin

The Application of Artificial Intelligence in Network Security Operation and Maintenance Services in the New Era

Bin Yuan

China Aerospace Changchun Control Technology Co., Ltd., Changchun, Jilin, 130000, China

Abstract

The Internet age, artificial intelligence technology, cloud computing technology in the network security operations services played a more and more important role, can realize the automation of network data security, automatic repair network vulnerabilities, intelligent prediction and prevention, network threat to ensure the safety of the network operation, reduce network information security threats, avoid network risk of economic losses to enterprise users, for the overall network security operational service level to provide enhanced technical support. This paper mainly analyzes the key points of the application of artificial intelligence in network security operation and maintenance services in the new era, so as to further improve the protection effect of network security operation and maintenance, reduce the occurrence probability of network security incidents, and optimize the service quality of network users.

Keywords

new era; AI; network security; operation maintenance services

新时代下人工智能在网络安全运维服务中的相关运用

袁彬

中国航发长春控制科技有限公司, 中国·吉林 长春 130000

摘要

互联网时代, 人工智能技术、云计算技术等网络安全运维服务中发挥了越来越重要的作用, 可以实现网络数据安全的自动化识别, 自动修复网漏洞, 对网络威胁智能预测和防范, 从而保障网络运行安全, 减少网络信息安全威胁, 避免网络风险对企业用户造成的经济损失, 为整体网络安全运维服务水平的提升提供强化的技术支撑。论文主要对新时代下人工智能在网络安全运维服务中的运用要点进行分析, 从而进一步提升网络安全运维防护效果, 减少网络安全事件的发生几率, 优化网络用户服务质量。

关键词

新时代; 人工智能; 网络安全; 运维服务

1 引言

随着互联网计算机技术的发展, 网络安全受到越来越复杂的威胁, 如网络漏洞、网络威胁、数据安全、隐私安全等, 加大了网络风险事件的发生概率, 对企业造成极大的经济损失。基于此, 需要对人工智能技术进行优化应用, 构建智能化网络安全运维管理系统, 进一步提升网络安全运维服务质量, 优化网络安全运行环境。

2 网络安全运维服务内容

2.1 健康检测

为了保障网络系统安全, 需要有效控制系统故障概率, 并延长系统使用时间, 在具体实施中需要对系统设备、业务

系统的健康状态进行全面性检测, 并对系统调配、流量耗用、系统运行状态等安全信息进行全方位搜集, 并实现用户系统安全检测的常态化, 以便对系统运转情况进行动态了解^[1]。

2.2 安全事件审计

随着互联网技术的发展, 网络安全事件逐渐增多, 且呈现复杂化、多样化趋势, 因此开发了大量的防火墙、IDS、VPN 等网络安全设备, 产生大量日志数据, 导致网络拥堵现象。基于此, 需要持续化搜集网络安全日志, 并优化安全审计, 完善安全评估。

2.3 网络行为审计

通过该服务的开展, 能够对网络用户群体、网络设备配置等情况进行全面审计, 尤其可以对网络应用历史进行掌握, 同时结合网络运行行为的各类数据, 实现网络环境的精准、有效优化, 全方位解读用户行为, 以此为依据提供动态滑动网络服务。在审计工作中, 一旦发现违规行为, 需要进

【作者简介】袁彬(1989-), 男, 中国吉林榆树人, 本科, 工程师, 从事网络安全研究。

行记录，方便事后追查和取证。

2.4 监控与分析

该内容主要是对网络资源安全进行管理和监控，并对海量网络数据进行深度挖掘，对不同事件的相关性进行构建，并采取针对性的安全监管措施，实现集中管控和智能化分析，实现统一监控，并对网络用户安全风险隐患进行预警处置，保障运维安全^[2]。

2.5 终端安全监控与策略完善

在现代化互联网技术的发展背景下，信息技术逐渐普及，网络用户越来越多，导致网络风险加大，加大了网络安全运维难度。当前企业网络安全监管不到位、病毒库滞后性等缺陷问题，会引起盗号、随意接入网络等安全风险，非常

不利于企业网络安全，甚至导致机密信息数据被盗取，对企业造成极大的经济损失。因此，要对人工智能技术进行优化应用，对用户终端安全进行实时监测，并优化安全监管策略，保障主体业务安全。

3 智能化网络安全运维架构

人工智能在网络安全运维服务中的应用架构如图1所示。其中，涉及基础层、技术层、应用层。基础层属于核心层，具有计算能力、数据资源等功能，能够进行数据获取、分析和处理；技术层涉及算法、模型和知识库，能够提取数据特征，并构建数据模型，实现模型评估与训练；应用层，即实现人工智能技术与网络安全服务的相互联合，提升运维服务质量^[3]。

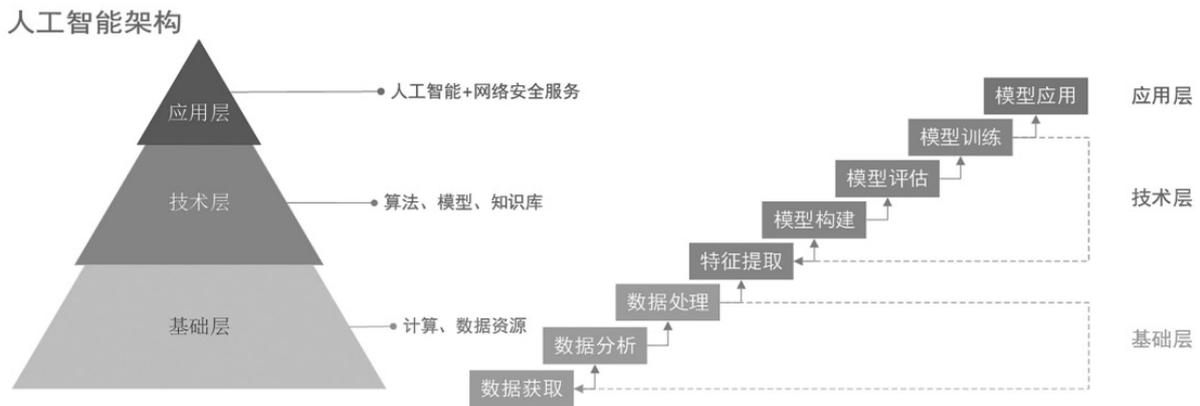


图1 网络安全运维中的人工智能架构

在人工智能技术支持下，能够利用云加载计算单元，实现中心化运维管理，进一步提升网络安全运维服务质量。其中，涉及事前、事中、事后三个环节，确保人工智能技术能够充分渗透到网络安全运维服务的全过程，构建智慧化安全运维体系，如表1所示。

表1 网络安全运维服务体系架构

网络安全运维阶段	具体事项
事前检测	配置基线、资产态势、漏洞检测、威胁分析
事中运维	安全监控、故障排查、数据备份、异常响应、安全态势
事后分析	日志审计、追踪溯源、安全分析、安全预测

①数据资产态势感知，随着互联网技术的发展，企业日常运营生产中形成的数据量呈现增长趋势，加大了数据处理压力。基于此，为了提升信息数据利用价值，要对人工智能技术进行优化应用，构建安全数据学习系统，对各类数据综合性采集，在此基础上，形成安全数据态势感知平台，以便对其动态监测，保障安全产品的优化布局。此外，还需要有效提升网络弹性功能，以便对网络安全数据进行动态化分析，精准分类安全报警信息，划分优先级，采取有效性、针对性的处理措施，构建多层次安全报告，保障企业安全管

理水平的提高。

②针对网络运维现状，需要对防火墙、IDS等技术进行优化应用，以便对APT高级持续威胁进行积极应对，避免对企业系统运行安全稳定造成威胁。此外，由于APT攻击存在一定的隐蔽性，难以进行有效性防控，需要利用人工智能技术，实现各类流量数据的有效性汇集分析，并深度挖掘，掌握攻击者线索。

③软件自动核查，针对企业开源代码的完整性风险，需要利用人工智能技术，全方位识别代码错误，有效控制代码漏洞数量，并进一步提升软硬件防御能力，强化企业对海量代码的分析能力。

④系统动态加固，针对基础设施面临的攻击问题，需要对人工智能技术进行优化应用，即通过自动识别技术，对系统、软件运行中出现的漏洞等进行全方位检测，自动生成补丁程序，并对其持续性修复，强化系统自动防御能力，强化系统加固效果^[4]。

⑤异常动态监测，随着网络技术的发展，网络攻击方式逐渐向多样化、复杂化、动态化、不确定性方向发展，加大了安全防护工作压力。因此，要利用人工智能技术的机器学习功能，在专业算法支持下，对历史数据进行收集分析，构建数据模型，对网络安全事件进行持续性训练、学习，以

便了解运维服务人员的数据需求，为其提供精准情报信息，以便提升响应速度，强化应急处置，减少网络攻击、异常事件造成的损失。

4 新时代人工智能在网络安全运维服务中的应用要点

4.1 自动化漏洞扫描与修复

在网络安全运维服务中，需要对人工智能技术进行优化应用，以便对网络漏洞进行自动化扫描和快速修复，从而强化网络安全防护能力。在具体应用中，需要在人工智能技术支持下，引进多元化的漏洞扫描工具、漏洞评估技术等，实现网络系统、应用程序漏洞的全方位检测，并加快对网络安全事件的响应速度。在机器学习、自然语言处理技术的辅助作用下，能够实现系统漏洞的定期扫描，并对海量漏洞数据全面性分析，并联合大数据技术深度挖掘数据内在关系，及时发现潜在的漏洞风险，并总结漏洞特征、规律，以便对系统漏洞进行精准预测，自动生成修复动作，减少漏洞风险问题的出现几率。自动化漏洞扫描与修复技术具有较强的持续学习和优化能力，能够对历史漏洞数据进行持续性汇总分析，进一步提升系统对漏洞的识别率，并完成自我进化，实现系统漏洞的高效性扫描和修复，保障企业网络安全。

4.2 智能化威胁检测与预警

在人工智能技术支持下，能够对网络流量、系统日志、威胁情报等海量数据进行实时监测，并对可能存在的网络威胁进行全面识别，分析潜在的网络攻击行为，为网络安全防范提供保障。在具体实施中，通过基于深度学习的安全威胁识别技术，对海量网络数据进行训练和学习，构建威胁识别模型，并对异常行为类型进行精准识别，实现潜在威胁的精准检测，并第一时间提取攻击源，形成封堵攻击源的指令，从而对网络攻击行为进行积极响应和防范。通过智能威胁检测和预测技术的应用，能够在机器学习、深度学习算法基础上，对海量数据进行汇集和分析，并提炼特征，利用模式识别、预测技术，对网络安全区域进行动态、精准分析，明确可能出现的攻击事件，保障企业网络安全，减少企业经济损失。此外，该技术还能够对社交媒体、电子邮件等网络数据进行全面性收集和识别，以便对威胁情报进行全方位掌握，强化网络安全监测，进度，进一步保障企业网络安全^[5]。此外，在自然语言处理技术的支持下，能够智能化分析文本数据，对可能出现的攻击行为进行精准识别，并对数据内在价值进行深度挖掘，以便对网络钓鱼、勒索软件、恶意软件等威胁进行全面性识别，提出针对性的防范方案。

4.3 隐私保护和数据安全

互联网具有开放性特征，且网络攻击手段逐渐升级，人们数据、隐私安全受到严重威胁，对安全事件响应方法提出了更高的要求。新时代，在网络安全运维服务中引入人工智能技术，如可以利用加密技术、差分隐私技术等进行融合应用，对网络安全风险进行快速识别，并智能化分析风险原因，提出针对性的处置措施，保障网络安全。在具体操作中，要利用深度学习技术，对安全事件进行自动识别，并对其进行精准分类，以便进行快速、针对性响应；利用强化学习技术，形成智能决策系统，同时实现安全事件的智能化分析，有效控制误报率；在差分隐私技术应用中，能够对网络海量数据进行汇总分析，并精准计算，从而保障用户隐私安全，避免出现信息泄露现象；加密技术的应用，可以利用安全多方计算技术，可以在保障用户私密数据安全的基础上，确保若干个参与者协同计算相应的函数，强化数据共享；同态加密技术的应用，能够在保障数据安全的基础上实现各类数据智能化计算，防护数据安全。在未来发展中，要不断优化算法和模型，保障安全事件的智能化、准确性响应，强化数据安全隐私保护技术，保障安全事件响应和处置工作的智能化开展。

5 结语

综上所述，为了提升企业网络安全运维服务水平，需要对人工智能技术进行优化应用，以便自动识别网络漏洞、智能检测网络威胁、加快安全事件响应速度，强化网络安全防护能力，真正实现网络安全运维管理工作的智能化、高效化，强化降本增效。

参考文献

- [1] 宋焱宏.一种基于人工智能的网络安全智能化运维技术解决方案[J].电脑知识与技术,2024,20(6):74-76.
- [2] 杨启航.人工智能在网络安全运维服务中的运用研究[C]//中国土木工程学会燃气分会,《煤气与热力》杂志社有限公司.中国燃气运营与安全研讨会(第十一届)暨中国土木工程学会燃气分会2021年学术年会论文集(下册).重庆燃气集团股份有限公司,2021:6.
- [3] 方雪琴,符方权,张嘉俊.浅析人工智能技术在电力企业网络安全运维中的应用[J].网络安全技术与应用,2021(3):104-106.
- [4] 金景峰.人工智能在网络安全运维服务中的运用研究[J].河南科技,2020,39(25):21-23.
- [5] 周利均.人工智能在网络安全运维服务中的应用[J].通信技术,2020,53(2):521-524.

Research on the Dynamic Scheduling Algorithm of Big Data Storage and Computing Resources for Cloud Computing

Chenyu Huo

Shanxi College of Applied Science and Technology, Taiyuan, Shanxi, 030000, China

Abstract

With the rapid development of cloud computing and big data technology, how to effectively manage and schedule resources to meet the needs of large-scale applications has become an important topic of current research. This study focuses on the dynamic scheduling of big data storage and computing resources in the cloud computing environment, and proposes a new dynamic scheduling algorithm. Firstly, this study establishes a resource scheduling model adapted to the requirements of big data processing according to the characteristics of the cloud computing environment. Then, we design a weight-based resource allocation strategy for the scheduling of big data storage and computing resources, and verify its effectiveness through simulation experiments. Finally, Proposed algorithm also takes into account load balancing and system utilization, and realizes efficient utilization and fair allocation of resources by introducing dynamic optimization strategies. The experimental results show that the proposed algorithm performs well in reducing task waiting time, improving system throughput, balancing load, and improving resource utilization.

Keywords

cloud computing; big data storage; compute resource scheduling; resource allocation strategy; task waiting time

面向云计算的大数据存储与计算资源动态调度算法研究

霍晨宇

山西应用科技学院, 中国·山西太原 030000

摘要

随着云计算和大数据技术的快速发展, 如何有效地管理和调度资源, 以满足大型应用的需求, 成为当前研究的重要课题。本研究聚焦于云计算环境下的大数据存储与计算资源的动态调度问题, 提出了一种新型的动态调度算法。首先, 本研究根据云计算环境的特性, 建立了一个适应于大数据处理需求的资源调度模型。其次, 针对大数据存储和计算资源的调度问题, 设计了一种基于权重的资源分配策略, 并通过模拟实验验证了其有效性。最后, Proposed algorithm也考虑到负载均衡和系统利用率的因素, 通过引入动态优化策略, 实现了资源的高效利用和公平分配。实验结果表明, 该算法在减少任务等待时间, 提高系统吞吐量, 平衡负载和提高资源利用率等方面均有较好的表现。

关键词

云计算; 大数据存储; 计算资源调度; 资源分配策略; 任务等待时间

1 引言

随着如今社会中大数据与云计算技术的高速发展, 我们正处于一个由过去的计算机硬件瓶颈转变为数据处理瓶颈的时代。在云计算环境中, 海量的数据和复杂的计算任务为资源管理和调度提出了极高的要求。尤其是对于维护和运行这样的大规模应用系统来说, 如何有效地动态分配和调度大数据存储和计算资源, 已经成为一个严重的挑战。因此, 如何设计和实施一种有效的资源动态调度算法, 以满足大数据处理需求, 对于提高云计算环境中的运算效率, 缩短任务的等待时间, 提高系统吞吐量以及实现资源的充分利用有着至关重要的意义。

2 大数据与云计算的资源调度需求

2.1 大数据的存储与计算需求

大数据的快速增长带来了海量的信息资源, 它的存储和计算需求是了解大数据和云计算资源调度的关键出发点^[1]。存储需求主要指的是大数据所需的存储空间和存储效率。大数据的海量性导致其存储空间需求迅速增大, 传统的数据存储方式难以满足大数据的存储要求, 急需一种能够实现高效、安全的存储方式。在海量数据存储的数据的存储效率也是大数据存储所要解决的关键问题。

计算需求主要包括数据处理速度和数据分析的准确性^[2]。大数据不仅仅是海量数据的存储, 更重要的是从这些数据中提取有价值的信息。如何快速、准确地分析处理大数据, 获得有用的信息, 是亟待解决的一大技术难题。大数据的异构

【作者简介】霍晨宇(2002-), 女, 中国山西长治人, 本科, 从事大数据管理与应用研究。

性特征（即数据类型的多样性），也要求其计算处理能力需具备处理多类型数据的能力。

大数据的存储与计算需求的满足，是大数据价值的有效实现，也是其在云计算环境下资源调度方案设计的重要基础。这种需求与大数据处理的实际应用紧密相关，只有满足了这些需求，才能真正实现大数据的价值，推动业务和科技的发展。

2.2 云计算环境的特性与资源调度需求

云计算环境具有动态扩展、按需付费、低历史成本和高服务质量等特点，为大数据存储与计算需求提供了有效支持。云计算环境中，资源主要包括计算资源、存储资源、网络资源等。这些资源的调度主要受到多维度数据特性、用户需求变化、服务等级要求和资源自身的性能特性等因素影响。云计算资源的实时、动态管理是决定云服务性能的关键。但基于大数据应用的复杂性和快速变化的特点，云计算环境中的资源调度也面临诸多挑战。

大数据应用具有繁多多变的计算模式与巨大的存储需求，这要求云计算环境中的资源调度策略必须具备强大的可扩展性和弹性，以便根据应用程序的具体需求动态调整。对于时间敏感的大数据应用，必须保证数据的实时处理与高利用率，这就要求云计算环境中的资源调度算法需具有高效执行和优秀的性能保障能力。用户需求的多样性以及任务的异构性也均对云计算资源的分配与调度提出了新的要求。在云计算环境中，资源调度策略的设计必须满足各类大数据应用的需求，保证资源的充分利用率，兼顾用户体验和服务质量。

3 新型动态调度算法的设计与实现

3.1 基于云计算环境适应性的资源调度模型建立

在云计算环境下，动态调度的首要步骤是建立一个适应于大数据处理需求的资源调度模型。要了解云计算环境的特性，包括其弹性、分布式、动态可调节性等。根据这些特性，资源调度模型应具备动态性、全局性和自适应性。

动态性是指模型能够实时响应各类资源的变化，对系统的工作负载进行动态调整。全局性则是指模型需要在整个云计算环境中，进行跨节点、跨地域的资源调整，以实现资源的最优配置和系统的最大性能。自适应性体现在模型能够根据应用的实际需求和资源的状态变化，智能地进行决策和调整。

在具备上述特性的基础上，模型中还需要引入管理策略。在资源管理策略中，引入了基于权重的分配策略，通过赋予各个应用不同的权重，可以有效地进行优先级的设定和资源的分配。考虑到云计算环境的动态变化，模型还需要实现实时响应和调整，对系统状态进行动态优化，确保系统的负载均衡和资源的完全利用。结合了这些策略后，资源调度模型就具备了优先级的描绘、动态的响应和自适应的调整，从而可以有效地满足大数据处理在云计算环境下的需求。

3.2 基于权重的资源分配策略设计

在新型的动态调度算法设计过程中，基于权重的资源分配策略起到了至关重要的作用。这种策略将云计算环境中的多种资源，如存储资源、计算资源等，根据其在大数据处理中的重要性给予不同的权重。

资源权重的设置主要分为两个步骤：进行资源特性分析，明确各类资源在大数据处理过程中的作用和重要性。例如，对于计算密集型的大数据任务，计算资源的重要性会相对较高，应给予较高的权重；而对于数据密集型的任务，存储资源的重要性就会比较突出。依据资源特性和任务需求进行权重分配，以此决定在资源分配中各类资源所占的比例。

基于权重的资源分配策略也具有较好的灵活性，能随着大数据处理任务特性的变化动态调整资源的权重分配，从而高效应对各种类型的大数据任务。该权重分配策略的设计，将极大增强云计算环境的资源调度能力，优化资源利用，降低系统资源的浪费，从而提升系统性能。

3.3 动态优化策略的引入与实现

动态优化策略在新型动态调度算法的设计与实现中发挥了至关重要的作用。为解决传统资源分配策略中遇到的负载均衡及系统利用率不高的问题，引入动态优化策略变得必不可少。在该策略中，采用动态调整权重的方式，实时响应和适应任务的需求和负载变化，保证了资源的高效利用和公平分配。

动态优化策略对系统中的任务进行实时监控，分析任务的执行状态、资源利用情况及系统的负载状况，并以此作为权重调整的依据。依据监控数据的变化，对资源分配策略中的权重参数进行动态调整，确保系统在任意时间都能以最合理的方式进行资源分配。

运用这种策略，无疑可以提高系统对存储和计算资源调度的灵活性和自适应性，使存储与计算资源的分配更加合理，负载更加均衡，系统吞吐量得以提高，任务等待时间得以减少，最终达到提高整个系统的效率和性能。这也提供了一种有效的资源管理策略，能为云计算系统的资源管理者在面对动态变化的存储和计算需求时，做出快速且准确的决策。

4 动态调度算法的效果评估与分析

4.1 实验设计与参数设置

实验设计的主要目标是验证所提出的动态调度算法在云计算环境下对大数据存储与计算资源管理的有效性。实验环境选择了开源的云计算平台 OpenStack，实验数据集则从实际业务系统中抽取^[9]。该系统具有大规模并行处理大数据的需求，数据量达到 PB 级别，任务类型复杂多变。

在参数设置上，主要考虑了五个因素：资源种类、资源数量、任务类型、任务数量和任务提交时间。其中资源种类分为计算资源和存储资源，资源数量从 500 个节点到

5000个节点不等,以满足不同规模的需求。任务类型包含计算任务和存储任务两种,计算任务按照CPU计算密集型,内存计算密集型,IO计算密集型三种类型随机产生,存储任务按照读写混合型,读密集型,写密集型三种类型随机产生。任务数量分别设定为1000、5000、10000、50000四个等级,以模拟不同规模的任务需求。任务提交时间则按照泊松分布来模拟真实环境中任务的提交模式。

在动态调度算法的参数设置上,主要涉及权重分配策略的参数和动态优化策略的参数。对于权重分配策略的参数,考虑了任务类型、任务大小、任务优先级等因素;对于动态优化策略的参数,考虑了系统负载、资源利用率等因素。这两套参数的设置旨在满足真实环境中对高效利用资源和公平分配资源的需求。

4.2 算法性能的评估指标

要准确评估所提出的动态调度算法性能,需要设计一系列的评估指标。其中,主要包括以下几个方面:

①任务等待时间:任务等待时间是指任务放入调度队列开始,到任务被分配执行资源结束的时间。可以通过对比不同算法下的任务等待时间,来验证所提出算法在延时优化方面的效果。

②系统吞吐量:系统吞吐量是指系统在单位时间内处理任务的数量。通过观察系统吞吐量,可以了解算法在提高系统处理能力方面的效果,这是衡量系统性能的一个重要指标。

③负载均衡:负载均衡可以从两个方面来衡量:一是各计算节点的负载差异,二是任务在计算资源上的分布情况。负载均衡是验证动态调度算法能否公平分配资源、优化系统性能的关键。

④资源利用率:资源利用率是指计算资源在一段时间内的使用效率。评估资源利用率可以验证算法在资源管理和分配中的优势,尤其是在大数据环境中,资源利用率是一个必要的考察指标。提高利用率意味着减少资源浪费,降低运行成本。

这四个指标的综合使用,可以全面、准确地评估动态调度算法的性能表现,并为进一步优化算法提供方向。

4.3 实验结果与分析

在对动态调度算法进行评估时,设定了多种实验场景并考察了任务等待时间,系统吞吐量,负载平衡以及资源利用率等关键因素。实验的具体过程中,通过模拟云计算环境中的各种复杂情况,对资源调度模型进行高压测试。

实验结果表明,与传统的资源调度算法相比,新型动态调度算法在任务等待时间上有显著的优势,能有效地减少任务的等待时间,从而提升了云系统的响应速度和用户体验。至于系统吞吐量方面,新型算法亦具有优异的性能,较为确切地平衡了系统的吞吐量,增强了系统处理大数据的能力。

跟负载平衡相关的实验结果显示,新型算法在均衡各计算节点负载方面明显超越传统算法。由于该算法的巧妙设计,能够实时地监控和调控各节点负载,从而实现高效的负载均衡,并且这种均衡的效果对于预防系统过载有着十分重要的意义。

在资源利用率方面,算法的优势同样很明显。它可以有效地增加资源的利用率,降低资源的浪费,从而最大化系统性能。

通过对上述各项因素的考察,证明了新型动态调度算法在大数据存储与计算资源的调度方面降低任务等待时间,提高系统吞吐量,平衡负载和提高资源利用率等方面具有显著优势,其结果对云计算系统的资源调度有着重要的指导意义。

5 结语

在我们的研究中,我们深入研究了大数据存储和计算资源在云计算环境中的动态调度问题,并提出了新的调度策略。我们建立了一个适应大数据处理需求的资源调度模型,并设计了一种基于权重的资源分配策略。通过实验,我们发现我们的策略可以有效地缩短等待时间,提高系统的处理能力,平衡负载,利用资源。这些对于负责云计算系统资源管理的人来说是很重要的。但是,我们的研究还有一些局限性,比如是否适应实际复杂的云计算环境,以及如何在不同的场景下公平高效地调度资源等问题。未来,我们会进一步研究不同任务和资源条件下的策略,同时考虑实际业务场景,使模型更实用、更普适。我们也会研究如何优化资源调度策略,提高系统效率和公平性,给云计算资源管理者提供更多的决策依据。

参考文献

- [1] 陈祖歌,毛冬,饶涵宇,等.面向物联网云计算的能源大数据存储优化算法[J].浙江电力,2023,42(8).
- [2] 许可.基于云计算的大数据存储安全[J].电子技术与软件工程,2019(12).
- [3] 石金梁,杨勇杰,吴玉亭.浅谈云计算任务资源调度[J].信息系统工程,2019,32(7).

Design and Implementation of a Distributed File System Based on Big Data Technology

Yanru Liu

Shanxi College of Applied Science and Technology, Taiyuan, Shanxi, 030000, China

Abstract

In the context of the big data era, traditional centralized file systems are no longer able to meet the needs of processing massive amounts of data, so researching and implementing distributed file systems is becoming increasingly necessary. The paper first analyzes the development background and current technical requirements of distributed file systems, and studies the core technologies involved in big data processing. Subsequently, using big data processing techniques such as HDFS and MapReduce as the main tools, a distributed file system for big data was designed. The system includes two major parts: distributed storage and parallel computing, and has optimized the access and data storage of common IoT devices. Experimental results have shown that compared to traditional centralized file systems, this distributed file system has higher efficiency and scalability in handling real-time analysis and queries in big data environments.

Keywords

big data technology; distributed file system; HDFS; MapReduce; Internet of Things device connection

基于大数据技术的分布式文件系统设计与实现

刘艳茹

山西应用科技学院, 中国·山西太原 030000

摘要

在大数据时代背景下,传统的集中式文件系统已无法满足处理海量数据的需求,因此研究及实现分布式文件系统显得越来越必要。论文首先对分布式文件系统的发展背景和当前技术要求进行分析,并对大数据处理过程中涉及的核心技术进行了研究。接着,运用大数据处理技术如HDFS和MapReduce作为主要工具,设计了一种面向大数据的分布式文件系统。系统包括了分布式存储和并行计算两大部分,并对常见的IoT设备接入和数据存储进行了优化。实验证明,相较于传统的集中式文件系统,该分布式文件系统在处理大数据环境下的实时分析和查询等任务具有更高的效率和扩展性。

关键词

大数据技术; 分布式文件系统; HDFS; MapReduce; 物联网设备连接

1 引言

大数据时代带来了巨大的数据存储和处理需求,传统的集中式文件系统难以应对现今日益增长的数据吞吐量,这使得研究和实现分布式文件的必要性凸显出来。分布式文件系统是一种能够实现数据的大规模分布和并行处理的存储系统,它可以将文件分布在网络上多台独立服务器的磁盘上,并对这些服务器进行协调,使得用户可以透明地访问尽管分布在远程磁盘上却像在本地磁盘上一样直观的文件。为了探寻有效面向大数据的分布式文件系统的设计与实现,本文基于HDFS和MapReduce等大数据处理技术展开深入研究,希望打造一种能够有效满足IoT设备接入,高效进行数据存储,并具备更强实时分析和查询能力的分布式文件系

统。在处理大规模数据时具有卓越的效率和扩展性,并实现数据的高可用和容错性。本研究的目标不仅是为分布式文件系统的实际应用提供有益的参考,也旨在推动大数据处理技术的发展与创新。

2 大数据时代的文件系统需求与挑战

2.1 海量数据的处理需求

在大数据时代背景下,海量数据的处理成为一项重要需求^[1]。数据规模的激增,带来了大量非结构化数据以及对读写速度和数据可靠性的严苛需求。对数据的读写和分析必须在短时间内完成,否则可能错过优秀的业务机会或者无法有效地解决问题。就实时性来说,实时的数据分析对于规模较大的组织和公司尤其重要,因为它们依赖实时分析来快速适应市场动态,满足客户需求,改进产品或者服务。

与此大规模数据的可靠存储和备份也是当前的一个重要挑战。为了保障数据的安全和完整,数据备份和容错机制

【作者简介】刘艳茹(2002-),女,中国山西岚县人,本科,从事大数据研究。

是不可或缺的。但是，由于数据规模的急剧增加，传统的备份方法已经无法满足需求，这导致了对新型的分布式存储的需求。这种分布式存储方法可以有效分担单一存储节点的压力，并且能够提供更高的数据冗余和可靠性^[2]。

在数据类型方面，所需处理的数据类型也日趋复杂。结构化数据只是海量数据中的一小部分，更多的数据如图像、语音、视频等非结构化数据占据了大部分比重。这种类型的数据需要复杂地处理和存储解决方案，而传统的数据库系统往往无法有效解决这类问题。

总的来说，在大数据背景下，随着数据规模的不断增加，这种处理的需求促使人们在技术层面寻求新的、更好的解决方案。

2.2 集中式文件系统的局限性

集中式文件系统在设计上存在固有的缺陷，这也是其无法适应大数据环境所导致的主要问题。在大量数据处理的过程中，集中式文件系统表现出明显的瓶颈和局限性。

首要问题便是集中式文件系统的弹性和扩展性。非分布式储存模式下的存储容量有限，难以应对数据量增长的需求，容量的扩展也涉及重大的硬件升级，对于企业来说，这既需要投入大量的成本，也带来操作的不便。与此相比，分布式文件系统则可以通过动态添加节点方便地扩展存储空间。

处理效率的问题也不能忽视。在集中式文件系统中，所有的数据操作都要通过集中的存储节点进行，极易造成磁盘 I/O 瓶颈，影响数据的读写效率。更为严重的是，在多用户访问的情况下，单一存储节点容易成为系统的瓶颈，导致整体处理性能下降。

再者，集中式文件系统在容错和高可用性方面存在致命的短板。一旦存储节点发生故障，整个文件系统就将出现服务中断的情况。尽管可以通过添加备份设备和使用 RAID 等技术来提升系统的可靠性，但在大数据的处理环境中，这些传统的容错和备份机制的效率和耗费问题依然突出。

3 大数据处理技术研究与分析

3.1 HDFS 的基本原理与应用

HDFS (Hadoop Distributed File System) 是一种分布式文件系统，专为大量数据处理需求所设计，满足分布式存储和并行计算的要求。HDFS 的基本原理是将数据分布式存储在机群中的各个节点上，通过每个节点的处理能力，实现数据的快速处理和访问。

HDFS 具有主从结构，即由一个单独的 NameNode (主节点) 和一批 DataNodes (数据节点) 组成。NameNode 管理文件系统的元数据，包括文件的目录结构、文件的属性以及每个文件所包含的数据块信息等。DataNode 负责存储和产生文件的数据载体——数据块。

在运行过程中，HDFS 的数据块默认大小为 64MB，

远大于传统文件系统中的数据块大小，这在大数据处理中能降低访问磁盘的次数，提高任务运行效率。数据块在 DataNode 上冗余存储，确保当某个 DataNode 出现故障时，不会造成数据丢失，保障了系统的容错性。

应用层面，HDFS 为上层大数据处理框架提供了稳定可靠的数据存储基础，也给用户提供了直接访问和操作数据的接口。借助 HDFS 强大的数据处理能力，大规模数据分布式处理任务的效率得以大幅提升。无论是数据挖掘、机器学习，或是实时分析查询等任务，HDFS 都能提供有效的支持。

3.2 MapReduce 的核心技术及优势

MapReduce 作为一种基于大数据的计算模型，以其便捷的并行化处理能力在大规模声明式编程中得到了广泛应用。MapReduce 计算模型主要包括 Map (映射) 和 Reduce (归约) 两种操作。在 Map 阶段，系统将输入数据切分为多个独立的数据块，并对每个数据块进行并行处理。具体操作为，输入一个数据对 (key/value)，经过用户自定义的 Map 函数处理之后，输出一组中间的 key/value 对。在 Reduce 阶段，系统会根据 key 进行数据的聚合，由用户自定义的 Reduce 函数对聚合的数据进行处理，输出最终结果。

MapReduce 的优势主要体现在以下几个方面：①易用性。用户只需要关注于编写自己业务的 Map 函数和 Reduce 函数，可实现任务的并行化处理，无需考虑数据分布、负载均衡等复杂的问题。②扩展性^[3]。由于 MapReduce 模型具有高度的并行化能力，可以轻易在成百上千台机器上进行分布式运行，以此快速处理大数据。③容错性。MapReduce 模型内部自带错误恢复机制，一旦某些节点发生故障，任务会自动在其他节点上重新执行，这样大大降低了由于节点出现故障而导致任务失败的可能性。④适应性。MapReduce 模型适应不同类型数据处理，既可以处理结构化数据，也可以处理半结构化和非结构化数据。大大丰富了其在大数据处理中的应用场景。

MapReduce 是一种十分适合进行大规模并行计算的理想选择，并且能够提高处理海量数据的效率，减少处理时间，达到了快速响应的目标。

3.3 IoT 设备接入与数据存储的问题与解决方法

在大数据环境下，物联网 (IoT) 设备产生的数据量日益庞大，正逐渐成为数据挖掘和分析的重要源头。传统的集中式文件系统在处理这些设备上产生的巨量数据时，普遍存在存储容量有限、数据处理速度慢、系统扩展性差等问题。

为解决这些问题，提出一种基于 HDFS 和 MapReduce 技术的分布式文件系统设计。此系统能有效接入和存储 IoT 设备产生的海量数据。由于 HDFS 具备出色的分布式存储能力，该系统可以将数据分散存储到多个节点上，不再受单一节点容量的限制，通过数据冗余提高数据的可靠性。采用数据分块存储的方式，不仅便于扩展存储空间，也有利于并行计算。

系统采用 MapReduce 并行计算模型，每一个数据块会被分配给一个计算节点进行并行处理，从而极大提升数据处理的效率。MapReduce 模型还可以实现动态任务调度和负载均衡，确保系统在处理海量数据时的稳定性和效率。

4 基于大数据技术的分布式文件系统设计与实现

4.1 系统的设计原则与架构

在设计基于大数据技术的分布式文件系统时，采取的核心设计原则是满足大数据环境下的存储和计算需求，以及保证高效性和容错性。

在系统架构上，分布式文件系统主要分为两个部分：分布式存储和并行计算。在数据存储部分，采用了 HDFS (Hadoop Distributed File System) 模型。HDFS 拥有卓越的横向扩展性，对大规模数据具有极高的存储和处理能力。通过数据块的概念，将文件分割成大小相等的块，并通过副本机制，使得失去单个或多个数据块时，系统能够从其他数据块中快速恢复数据，具有很高的容错性。

在并行计算部分，引入了 MapReduce 计算模型，有效解决了海量数据的分布式处理问题。MapReduce 将大规模计算问题拆分成多个子任务，在分布式计算环境中并行处理，大大提高了计算效率。

除此之外，为了进一步优化系统的功能，考虑到 IoT 设备数据接入的特殊性，设计了优化的接口和数据存储方式，能够快速高效地接入和处理 IoT 设备数据。

总体上，本分布式文件系统的设计以高效、容错、可扩展为核心，能够满足大数据环境下的复杂需求，具有较高的实际应用价值。

4.2 分布式存储与并行计算的实现

在基于大数据技术的分布式文件系统设计与实现中，分布式存储与并行计算的实现是一个关键环节。系统整体设计采用主从结构，以 HDFS 作为分布式文件系统的主要存储框架，它具备容错性、高并发性和可扩展性的优点，且能有效支持海量数据的存储需求。对于存储数据的物理节点，采用数据块的方式进行数据切分，并进行副本备份，增加系统的容错性和数据的可用性。

在并行计算方面，系统主要选用 MapReduce 模型，它将计算任务分解成数个小的子任务，分发到各个节点上执行，再将计算结果进行汇总，实现了数据的分布式处理。为了配合这一计算模型，文件系统在存储数据时，会考虑到数据局部性的问题，尽量将需要共同处理的数据保存在同一个

节点上，减少网络间的数据传输，进一步提高处理速度。

针对 IoT 设备接入和数据存储的优化也同样重要。系统通过实现数据的多级缓存和调度策略，降低了 IoT 设备的数据访问延时，提高数据的存取效率。为了适应不同设备和应用的需求，系统还设计了一套灵活的 API 接口，以支持不同的数据模型和查询方式，满足个性化需求。

4.3 分布式文件系统的效率和扩展性测试

对于基于大数据技术的分布式文件系统，效率和扩展性是其绩效成败的关键。在实践阶段，对系统的模拟运行和实际操作进行了一系列的测试。

在效率方面，使用真实环境的海量数据对系统进行了压力测试。通过比较该分布式文件系统与传统的集中式文件系统在处理相同的工作负载时的时间花费，得出该系统具有优越的数据处理效率。尤其在大规模数据实时分析和查询方面，分布式文件系统凭借其并行计算的优势，表现出了更强的处理能力。

在扩展性方面，实验者模拟了各种扩展情景，包括节点数翻倍、数据量翻倍以及业务冗余等，以测试系统在不同环境下的适应能力。结果显示，随着节点数和数据量的增加，系统的表现始终保持稳定。尤其是在面临大规模数据增长时，系统采用了分布式存储，实现了数据量的线性扩展。

综合以上测试结果，基于大数据技术的分布式文件系统在效率、扩展性以及容错性等方面，相比传统的集中式文件系统具有较大优势，能够更好地满足大数据时代的需求。

5 结语

本研究基于大数据时代的背景，深入研究了分布式文件系统，并且实现了一种新的分布式文件系统。例如，如何在大规模并行处理中保证数据的一致性和完整性，以及如何适应不断变化的大数据环境，都是需要进一步探讨的问题。同时，研究也指出未来可能的发展方向，包括对更多形式的数据进行支持，对新的并行计算模型进行研究，以及提高系统的稳定性和可用性等。总而言之，研究为面向大数据处理的文件系统的设计提供了新思路和实践经验，对于深化分布式文件系统的研究和实践有着积极的推动作用。

参考文献

- [1] 陈行滨,王周,郑飘飘,等.基于分布式文件系统电力大数据存储实现[J].粘接,2022(6).
- [2] 刘军,冷芳玲,李世奇,等.基于HDFS的分布式文件系统[J].东北大学学报:自然科学版,2019,40(6).
- [3] 张海涛,张文娟.基于大数据的分布式文件系统技术研究[J].电子测试,2019,30(4).

Reflection on Software and Hardware Development Strategy Based on Embedded Design

Dejuan Chen

Shandong Xintong Electronics Co., Ltd., Zibo, Shandong, 255000, China

Abstract

With the continuous improvement of science and technology level, the embedded system is more and more widely used in all aspects of social life. Embedded systems refer to electronic products that integrate computers and controllers and can automatically perform tasks during operation. The development of hardware and software in the embedded system is the most important part of the development of the whole system. The selection of appropriate development technology strategies and the optimization of the overall design can give full play to the performance advantages of the embedded system and meet the application requirements of various fields. Based on this, the research work of this paper, briefly summarizes the characteristics of the embedded system, based on the embedded design software and hardware development strategy and related design methods, in order to provide reference and help for relevant personnel.

Keywords

embedded design; hardware and software; development strategy

基于嵌入式设计的软硬件开发策略思考

陈德娟

山东信通电子股份有限公司, 中国·山东 淄博 255000

摘要

随着科学技术水平不断提升, 嵌入式系统在社会生活的方方面面的应用越来越广泛。嵌入式系统指的是集成了计算机和控制器的电子产品, 在运行的过程中可以自动执行任务。嵌入式系统中的软硬件开发是整个系统开发中的最重要的部分, 选择合适的开发技术策略, 优化整体设计, 可以充分发挥嵌入式系统的各项性能优势, 满足各领域的应用要求。基于此, 开展本文的研究工作, 简单概述嵌入式系统的特点分析, 基于嵌入式设计的软硬件开发策略和相关的设计方法, 以为相关人员提供参考和帮助。

关键词

嵌入式设计; 软硬件; 开发策略

1 引言

嵌入式系统的开发和应用中软硬件的开发尤为重要, 而最为关键的是要做好软硬件功能的划分工作, 因此要选择合适的开发策略, 优化整体设计, 从而确保嵌入式系统的正常稳定运转, 取得显著成效。ARM、FPGA 等为基础的嵌入式系统应用比较广泛, 基于此, 开展分析工作, 明确具体的开发策略。可以选择去中心化设计思路, 进行模块功能设计, 完成软件硬件协同设计工作, 发挥模块优势, 确保多模块独立运行和密切联系, 降低设计成本, 缩短开发周期。

2 嵌入式系统的概述

嵌入式系统以信息技术作为基础核心, 依托于计算机

系统进行相关功能的研发, 满足市场的相关需求, 在系统架构方面主要分为硬件与软件部分(图1)。此类系统可以独立运作。其中硬件包括通信模块、存储器、信号处理器等, 软件仅仅是操作系统和运行环境。在开发工作中要选择合适策略, 优化设计流程。可以通过软硬件协同设计, 模块化处理, 升级现有系统。嵌入式系统在很多产业中得到广泛应用, 如工业自动化、国防、航天等领域。嵌入式系统迅速走上了独立发展的单片机道路, 向微型控制与片上系统发展。嵌入式系统具有嵌入性、专用性与计算机三个基本特点。嵌入性指的是计算机嵌入到对象体系中, 实现对象体系的智能控制。计算机是对象系统智能化控制的根本保证、专用性指的是要满足对象控制要求下的软硬件裁剪性, 进行软硬件的优化配置。因此, 在设计开发工作中, 相关人员遵循以上原则, 满足应用要求。

【作者简介】陈德娟(1988-), 女, 中国山东济宁人, 硕士, 从事仪器仪表研究。

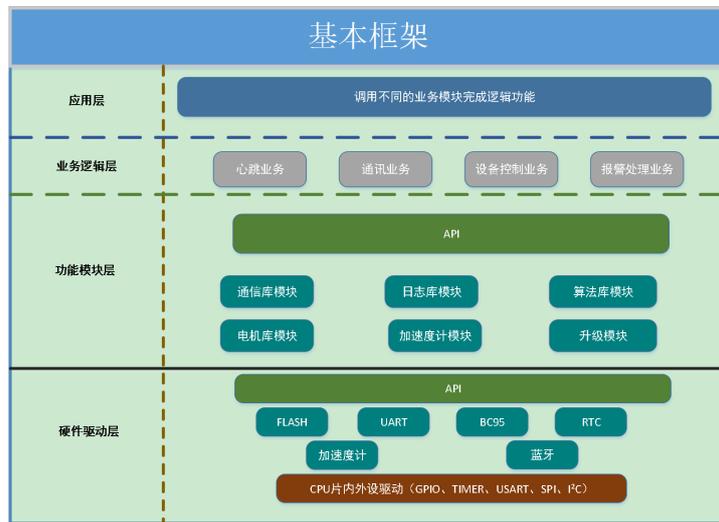


图1 嵌入式系统的基本框架

3 基于嵌入式设计的硬件开发

在嵌入式系统的硬件开发工作中，主要包括嵌入式微处理器、存储器、各类接口等。在此基础上添加电源电路、时钟电路和存储器电路构成了嵌入式核心控制模块。

3.1 处理器

处理器是嵌入式系统中的管理控制中心，开展对整个系统的控制工作，调试各项任务，加强管理工作。以CS89712嵌入式处理器为例，使用双电源供电内核电压值控制在2.5V左右，可以进行多种时钟设置，外设接口丰富，同时支持4个内存块与内存接口开展编程操作。以FPGA为基础的嵌入式系统中，可以选择EPCS1作为存储器，为系统的开发提供支持。完成开发环境编译后，借助于接口将执行文件下载至开发板中的flash上，为嵌入式操作系统提供支持。

3.2 存储器

存储器用来存放和执行代码，包括Cache储存和辅助存储器等。Cache是一种容量小速度快的存储器阵列，主要的目的是减少存储器给微处理器内核造成的访问瓶颈，可以存放微处理器，使用最多的程序代码和数据。在应用过程中，微处理器可以从其中读取数据，不需要从储存中获取，可以改善系统的整体性能，提高数据的传输速率^[1]。储存是微处理器直接访问的寄存器用来存放系统和用户的程序与数据，容量小，速度快。辅助存储器用来存放大的数据量的程序代码或者信息，读取速度相对较慢，不过可以长期存储用户的信息。全世界嵌入式微处理器多种多样体系结构有30多个系列，如ARM、MIPS等。通用设备接口和I/O接口设计中，包括A/D D/A、I/O、RS-232 USB、串行接口、电源接口等。硬件系统构成如图2所示。

3.3 中间层的开发

软件层与硬件层之间为中间层，也指的是硬件抽象层或板级支持包，能够使上层的软件与底层硬件分离开来。中间层包括底层硬件的初始化数据的输入与输出、操作硬件设

备的配置功能。

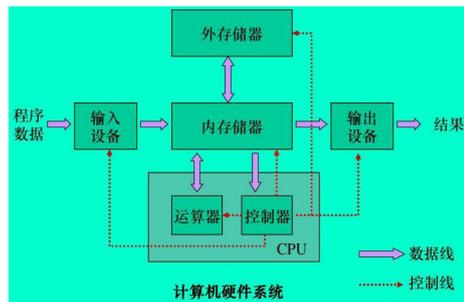


图2 硬件系统构成

4 基于嵌入式设计的软件开发

4.1 软件的应用开发

嵌入式软件主要包括操作系统、应用系统和底层驱动。操作系统简称OS，承担系统的任务，处理进程调度的各项功能。在操作系统的设计工作中，要注重实时多功能操作系统的开发设计工作。包括嵌入式GUI系统和电源管理。随着不断地开发与发展，促进了嵌入式系统的完善性，也增强了嵌入式软件的流动性。具有开发性能高、应用广泛、实时功能强的特点，为嵌入式系统提供良好的服务^[2]。例如，开发MicroC/OS-11实时操作系统时，建立相应的软件工程开发程序，将其设计在嵌入式系统中使其稳定运行。编写两个互相调用的任务，可以实现整体系统运行情况的测试工作。增设两个任务之间互相调用的代码，开展执行测试，可以达到相互调用效果，完成操作系统的设计运行。

GUI指的是利用图形方式显示的计算机用户操作界面，嵌入式GUI技术的不断发展，代表着计算机技术的进步，可以减少用户的认知力，负担使用用户界面的显示，使用更加的人性化。可以选择嵌入式LinuxGUI，可以应用于手机数控平台家电等的运行平台中。具有性能高、可靠性高、轻便配置性高的特点。

4.2 相关技术应用

4.2.1 IP 构件技术

IP 构建技术指的是常见组件开发模型。IP 构建作为软件开发的基础开发平台，应用越来越广泛，在设计工作中应用 IP 和实现对芯片系统核心功能的整合工作。将多项的附属功能集成到单一的芯片内部，发挥集成化的优势，完成构件开发工作，可以减少整体的开发周期，升级设计功能，提高工作效率^[3]。

4.2.2 交叉开发技术

在开发嵌入式系统的软件时，目标机与宿主机之间进行任务的交互和资源的分配，宿主机要注重硬件资源的定制编译，从而开展工作模块的开发工作。不过最终的嵌入式软件实际运行平台是目标级，因此可以使用交叉开发技术，开展软件开发工作。将宿主机运行应用程序转变为二进制代码，可以实现在目标机上的运行，利用交叉编译完成整体的编译工作，形成目标文件，在目标机上运行使用。

5 基于嵌入式设计的软硬件开发策略

5.1 软硬件协同设计

SOC 是芯片级系统，也是嵌入式系统中的片上系统，是多个具有特定功能的集成电路组合在一块模块上形成的系统和产品，包含完整的硬件系统和承载的嵌入式软件。SOC 由两部分组成，一部分是固件在固件的基础上开发，只需要调用底层硬件、驱动程序、文件管理系统等部件的 API 函数，既可以开发出一个稳定可靠的产品。另一部分是硬件模块化，它具有复杂的高性能的嵌入式，处理的最小系统和特定功能，可以被直接调用。SOC 核心硬件设计难度大。软件的比重大。在单个模块上集成了更多配套的电路，节省了集成电路的面积和整体的成本，也确保各器件之间信息传输更加快速。在设计工作中引入了模块化设计思路，实现全系统的模块化设计，用户在使用时只需要更换各部分的模块和嵌入结构就能满足要求，它的优点是系统能够更加接近理想系统的状态^[4]。在开发工作中要选择合适技术，投入大量的资金和人才，优化整体设计工作，满足市场的发展需求。

5.2 去中心化式的嵌入式设计的软硬件开发

开展嵌入式系统的软硬件开发工作时，需要明确划分软硬件所需要实现的各项功能，从而优化设计对应的开发流程，选择合适的技术方法。在具体的设计工作中，嵌入式芯片的开发是与软件结合应用的，因此软硬件的划分界限比较模糊，因此可以采用去中心化式的设计方法，对整个开发模式进行重构设计一套全新的流程思路。在去中心化的支持下，可以明确嵌入式系统的各项功能，分散模块，以模块为中心开展设计工作。将集中化模式转为资源均衡分散模式。在驱动信息化的支持下，CPU 不再是嵌入式系统整体的处理单元，而是作为一个桥梁，连接多个独立模块，形成多处理器模式^[5]。CPU 上承载的业务分布在多个 CPU 中进行流转，可以定制化芯片，处理一些复杂的数据和大量的重复逻辑业务，均衡负载。同时还能实现模块化的处理划分，不同

模块确保能够独立运行，在该部分模块固化后可以实现高度聚合，通过接口与外部模块连接，形成一个完整的系统。在控制与显示工作中，只通过接口程序接收其他模块发送的控制信息和显示信息。

5.3 基于组合算法的软硬件划分

通过遗传算法与粒子群优化算法的结合应用，可以找寻嵌入式系统的最佳划分方案。发挥智能化算法的优势，提升嵌入式系统软件开发的效率。将嵌入式系统进行软硬件细分与节点相结合开展硬件与软件的迭代处理，可以克服两种算法的优缺点，寻找嵌入式系统的最优分布。嵌入式软硬件的数字模型可以解决遗传算法的相关问题，粒子群算法也找到了最优解，结合应用，改进硬件与软件的整体设计。个体的初始种群是随机的，进行个体的优化选择，可以进行交叉与变异操作，形成新的组合。进行遗传算法，形成粒子群，确定粒子群的最佳算法。优化组合的最佳反编码，从而选择最佳软硬件划分算法，完成嵌入式系统的开发与运行。

5.4 嵌入式系统的初始化

嵌入式系统初始化，按照从硬件到软件，包括片级初始化、板级初始化和系统级初始化。首先完成嵌入式微处理器的初始化，包括微处理器的核心寄存器和控制寄存器、相关的工作模式和局部总线模式。从默认状态逐步设置为系统要求的工作状态，实现纯硬件的初始化过程。其次，完成其他硬件设备的初始化设置。调整软件的数据结构与参数为整体的初始化运行，建立硬件和软件环境。最后，BSP 对嵌入式微处理器的控制权转移到操作系统中，在操作系统完成剩下的初始化操作工作。加载设备驱动程序，建立系统内存区，初始化其他的系统软件模块，最后创建应用程序环境，将控制权交给应用程序入口。

6 结语

综上所述，开展嵌入式系统的软硬件开发工作，要具备软件硬件协同化设计和去中心化设计的思路，实现复杂业务逻辑模块，划分各功能模块独立运行，提高了信息传输的效率，有效规避故障，可以满足多功能需求，随时替换各类模块。在开发设计工作中，要注重先进模式的合理应用，对整个系统进行初始化设置，设置合适的参数，确保整个系统运行的稳定性，为各行业应用提供支持。

参考文献

- [1] 杨熙.基于嵌入式设计的软硬件开发策略[J].网络安全和信息化, 2022(11):92-94.
- [2] 赵明.基于嵌入式设计的软硬件开发分析[J].电脑知识与技术, 2021,17(12):208-209.
- [3] 门殿卿.基于嵌入式设计的软硬件开发分析[J].通信电源技术, 2022,39(11):68-70.
- [4] 牟晋力.嵌入式系统中软硬件协同设计技术的运用策略[J].通信电源技术,2022,39(15):161-163.
- [5] 李攀.基于嵌入式设计的软硬件开发分析[J].电脑高手(电子刊), 2020,2(2):1544.

Application Status and Improvement Strategy of Network Data Mining in Business Decision-making

Shuai Wang

School of Software, Sichuan University, Chengdu, Sichuan, 610065, China

Abstract

With the rapid progress of big data technology, the Internet has penetrated into various industries. Network data mining technology can not only significantly improve the efficiency of data use, but also provide a solid scientific foundation for the sustainable development of individuals and enterprises. At present, in the era of big data, enterprises have realized the important impact of data analysis and mining on the whole business decision, and take it as one of the future development trends. In the context of the era of big data, analyzing and studying the application of network data mining in business decision-making becomes particularly important. This requires the relevant staff not only to master and skillfully use this technology, but also need to improve the work efficiency at the same time, reduce the work required capital investment, so as to lay a solid foundation for the sustainable and stable development of the society. Based on this, this paper focuses on the application status and improvement strategy of network data mining in business decision-making.

Keywords

network data mining; business decision making; application status; improvement strategy

网络数据挖掘在商业决策中的应用现状及改善策略

王帅

四川大学软件学院, 中国·四川成都 610065

摘要

伴随着大数据技术的飞速进步, 互联网已经渗透到各个行业中。网络数据挖掘技术不仅能够显著提高数据的使用效率, 还能为个人和企业的持续发展提供坚实的科学基础。目前在大数据时代下, 企业已经意识到了数据分析与挖掘对整个商业决策的重要影响, 并且将其作为未来的发展趋势之一。在大数据时代的背景下, 分析和研究网络数据挖掘在商业决策中的应用变得尤为重要。这要求相关的工作人员不仅要掌握和熟练运用这项技术, 还需要在提高工作效率的同时, 减少工作所需的资金投入, 从而为社会的持续和稳定发展奠定坚实的基础。基于此, 论文针对网络数据挖掘在商业决策中的应用现状及改善策略进行重点分析。

关键词

网络数据挖掘; 商业决策; 应用现状; 改善策略

1 网络数据挖掘概念分析

作为一种现代化的技术, 网络数据挖掘技术在各个行业领域体现出较高的应用价值, 能够在第一时间掌握相关产品使用情况, 对海量的数据信息进行优化, 以此提高数据信息的利用效率, 为商业重要决策提供可靠的参考依据。根据以往数据挖掘技术的发展情况来看, 这一技术已经发展到能够从模糊、复杂的数据中找出有用信息, 所以在技术方面提出的要求较高, 想要发挥这一技术的作用, 就必须重视对互联网技术的运用。

网络数据挖掘的核心技术是一种搜索活动, 它主要依

赖特定的算法来进行探索, 并利用特定的工具, 从众多的数据中筛选出对搜索者有价值的信息资源。在数据挖掘中, 需要先建立模型, 然后对模型进行处理。这一流程的初始数据被划分为模型构建样本和分析样本两个部分。在数据分析前, 建模样本先与分析样本相结合。经过数据的预处理, 建模的样本将转化为经过预处理的专家样本。分析的样本在经过特征筛选之后, 会被转化为经过预处理的分析样本。对于预处理后的特征采用支持向量机进行分类。通过对专家样本和分析样本进行结果评估和建模规则的筛选, 最终生成了商业决策所需的数据挖掘结果。

2 网络数据挖掘在商业决策中的应用意义

随着互联网和信息技术不断发展, 网络数据挖掘技术成为商业发展关注的重点内容之一, 网络数据挖掘技术在提

【作者简介】王帅(2003-), 男, 中国河北沧州人, 在读本科生, 从事信息系统与数据分析研究。

高商业决策数据信息挖掘的便捷性上起到了不可或缺的角色。数据挖掘是从大量数据中提取出有用信息、模式和规律的过程。它涉及统计学、机器学习、数据库技术和可视化等多个领域。数据挖掘的目标是通过分析数据，发现数据背后的潜在模式和关联，以支持决策制定、预测未来趋势、发现新的知识和优化业务流程。数据挖掘的应用范围广泛，包括市场营销、金融风险、医疗诊断、社交网络分析等领域，成为了企业和组织中的重要工具。

网络数据挖掘在传统数据分析方法基础上做了重大扩展和提升，在分析复杂数据集方面展现出了巨大优势，尤其在互联网及人工智能时代，商业决策工作其应用范围和深度正在不断拓展。计算机技术进步极大地促进了数据挖掘领域发展，使数据分析更为准确高效。随着技术发展，计算机不仅增强了处理大规模数据集的能力，还通过精准的算法分析，保障数据分析结果的真实可靠。且该技术的应用范围也遍及各个行业，进而使得数据挖掘的作用变得日益重要，支持各种数据的统计学分析，为不同学科提供了强大的数据支持。在信息量急剧增长时代，商业决策过程中网络数据挖掘技术应用对于处理海量数据变得尤为关键，不仅可以对收集到的数据进行精确统计分析，还能针对特定数据进行灵活处理，从而使得数据被有效收集和存储，并深入分析与利用。信息化社会迅猛发展的大背景之下，大数据技术、云计算技术和物联网技术纷纷涌现，它们在日常生活和生活中产生了大量的数据和资料，从而推动社会步入大数据的新时代。

3 网络数据挖掘在商业决策中的应用方法

3.1 大数据存储技术和架构

网络数据挖掘过程主要是从大量数据中提取有价值的信息，此过程不仅依赖于多种数据采集方法，如神经网络、决策树、遗传算法等，还包括从问题定义到最终结果评估的全面步骤。随着互联网技术的快速发展，越来越多先进的信息技术应用到了互联网的方方面面，尤其是网络数据挖掘技术能够对数据信息进行深入分析和计算，进而针对商业决策中产生的各项数据进行科学探究，以更好地服务于商业决策工作。

分布式文件系统是一种能够在多台服务器上分布式存储数据的文件系统，如 HDFS 和 GFS 等，它们通过数据分片和冗余备份实现了数据的高效存储和可靠性保障。而 NoSQL 数据库则是一种非关系型数据库，如 MongoDB 和 Cassandra 等，它们通过分布式存储和水平扩展的方式，实现了对海量数据的高效管理和查询。云存储则是一种将数据存储在云端的方式，如 AWS S3 和 Azure Blob Storage 等，它们通过弹性的存储空间和按需付费的方式，为商业决策提供了高效的数据存储和管理服务。大数据存储架构主要包括数据采集、数据存储和数据管理三个环节。数据采集是将分布在各个数据源中的数据进行抽取和汇总，实现了数据的集中

化管理。数据存储是将采集到的数据进行存储和管理，保证了数据的安全和可靠。数据管理则是对存储的数据进行分析和查询，为商业决策提供了高效的数据检索和分析功能。

3.2 网络数据挖掘分析算法

随着科技的不断进步和信息化的快速发展，网络数据挖掘技术已成为当今社会的重要组成部分。未来，网络数据挖掘技术将继续呈现出快速发展的趋势，数据量将持续增长，数据类型将更加多样化，数据处理技术将更加智能化，而边缘计算技术也将日益兴起，成为大数据处理的重要方式之一。网络数据挖掘分析算法种类繁多，常用的包括聚类分析、关联规则挖掘、分类分析和预测分析等。聚类分析通过对数据进行分组，发现其中的内在结构和模式，常用的算法包括 K 均值和层次聚类。关联规则挖掘则是寻找数据中的关联关系和规律，用于市场分析和推荐系统等。分类分析是对数据进行分类或标记，商业决策网络数据挖掘常用的算法包括决策树、支持向量机和朴素贝叶斯等。预测分析则是基于历史数据进行趋势预测和未来事件预测，其中包括线性回归、时间序列分析和神经网络等算法。

3.3 商业决策数据预测分析

考虑到商业决策中涉及的数据种类较多，影响因素也相对较多，因此需要整合多源数据挖掘技术提高数据预测的精准性，常见的技术方法包括 ARMA 时间序列分析、人工神经网络等，技术人员可综合常见技术方法，在油田生产开发的历史数据基础上，构建 ARMA 时间序列分析模型以及人工神经网络模型，并调节好对应模型的参数，并将预测结果与实际数据进行比较，从而获取到最优的数据预测模型。之后，再将相关参数数据导入到数据预测模型中，获得最终预测结果，以为商业决策提供有效的数据支撑。

4 网络数据挖掘在商业决策中的应用现状

随着大数据时代的来临，网络数据挖掘在各个领域得到了广泛应用。然而，随着数据量的爆炸式增长，网络数据挖掘在商业决策中的应用也面临着前所未有的挑战。

①大数据时代带来了大量的数据资源，但同时也带来了数据安全和隐私保护的挑战。网络数据挖掘过程中需要处理大量的敏感信息，如个人身份信息、金融交易记录等，一旦泄露或被滥用，将对个人隐私和企业利益造成严重威胁。因此，如何在保证网络数据挖掘质量的同时，确保数据的安全和隐私保护，是网络数据挖掘技术面临的重要挑战之一。

②大数据时代的数据量庞大且复杂，如何高效地处理这些数据并保证挖掘结果的精度，是数据挖掘技术的另一个挑战。传统的数据处理方法和算法在处理大规模数据时可能会遇到性能瓶颈，无法满足实时性和准确性的要求。因此，需要不断优化数据处理算法和工具，提高处理效率和精度，以满足不同应用场景的需求。

③大数据时代的数据来源多样且复杂，数据的准确性

和质量参差不齐,这给数据挖掘带来了很大的困难。在进行数据挖掘之前,需要对数据进行清洗、去重、转换等预处理工作,以提高数据的质量和一致性。然而,预处理过程可能会消耗大量的时间和资源,因此如何在保证数据质量的前提下提高预处理的效率,也是数据挖掘技术面临的挑战之一。

④随着大数据时代的不断发展,传统的网络数据挖掘算法可能无法满足新的应用需求。因此,如何对算法进行优化和创新,以适应不断变化的应用场景和数据特征,是数据挖掘技术面临的又一挑战。同时,新的算法和技术的研发也需要大量的时间和资源投入。

综上所述,网络数据挖掘技术在大数据时代面临着数据安全和隐私保护、数据处理效率和精度、数据质量和预处理以及算法的优化和创新等多方面的挑战。为了应对这些挑战,商业决策过程中需要不断加强技术研发和创新,提高数据处理能力和算法精度,同时加强网络数据安全和隐私保护措施的应用,以满足不断变化的市场需求和应用场景。

5 网络数据挖掘在商业决策中的改善策略

5.1 利用网络数据挖掘技术,提升商业决策水平

商业决策系统在功能与架构确定后,需要通过有效措施进行系统优化设计。其中,系统数据库设计和系统数据分析决策生成为核心内容,系统数据库需要根据商业决策需求进行数据的获取与整理,形成样本数据,根据分类规则获得预测结果,最终为商业决策系统的运行提供可靠支持。系统生产开发数据库建设中,需要实时获取生产数据信息,并将获取到的信息存储到数据库中,通过数据整合分析为后续生产开发决策活动提供数据支撑。通过基础数据信息整理后获得样本数据可体现出商业决策工作的表征特征与决策属性,此类属性需存储在专门的样本数据中,用于后续的商业决策应用。此外,数据库中的分类规则包含大量的数据范围标准,其主要是各类数据获取的标准依据,能够为决策工作提供辅助参考,按照多元分类规则,可有效提取出样本数据中的表征特征和决策属性,提高数据挖掘分析效率。系统中的预测结果数据库,用于存储阶段性的数据预测结果内容,此项数据具有明确时间标定,同时会根据其他基础数据的更新情况对商业的运行情况进行阶段性分析,并且会将更新获得预测性数据及时存储供决策工作使用,决策结果数据库中包含多种决策方案内容、预测结果描述信息和预测结果范围标定,通过此项数据能够清晰地了解商业的决策情况。

5.2 构建网络数据挖掘模型,提升商业决策效率

网络数据挖掘是从大量数据中发现潜在的、先前未知的、可理解的、可适用的、最终可用的知识的过程。它涵盖了数据预处理、模式发现、模式评估和知识表示等过程,通过统计学、机器学习、人工智能等技术手段,探索数据之间

的关联、趋势和模式,为决策提供支持。数据挖掘的基础在于对数据的分析和挖掘,通过挖掘数据背后的价值信息,为企业决策和业务发展提供重要参考。人工智能(AI)驱动的网络数据挖掘领域,网络数据驱动模型的构建具有重要意义,基于深度学习的数据模型代表了高度复杂且灵活的数据处理方式,能够深入挖掘数据内在规律。同时针对大规模数据集的处理与优化成为关键,需要创新性的算法和技术以应对数据量的爆炸式增长,实时商业决策数据处理与应用则强调了即时性与实用性,能够带来更快速的决策和响应能力,但也对系统的稳定性和效率提出了更高要求,这些方面的发展为AI与大数据结合带来了前所未有的机遇与挑战。AI与大数据的融合涌现出一系列新技术,其中智能数据分析技术是引人注目的一环,这项技术利用AI算法处理数据,从而提取信息和趋势,帮助商业发展做出更明智的决策。跨领域数据整合与交叉应用则打破了传统行业的壁垒,不同领域的数据相互整合、交叉应用,为创新和发展提供了更广阔的空间,数据模式识别和预测能力的提升是AI与大数据结合的必然趋势,通过深度学习和神经网络技术,系统能够更精准地识别数据模式,并做出更可靠的预测,这在商业决策和科学研究中发挥重要作用,这些新技术的出现为大数据时代的发展提供了前所未有的机遇性。

6 结语

综上所述大数据的时代背景下,网络数据挖掘技术的实施与创新显得尤为关键,商业决策对大数据挖掘技术的高效应用和深入分析是不能被忽视的。在实际应用过程当中,数据处理与挖掘是一项较为复杂的工作,其涉及到了多个学科领域。技术专家们需要从多个视角去深入理解网络数据挖掘技术的真正含义,并对不同行业和模块中的大数据技术应用需求进行细致分析。通过全方位的调整和优化,可以更有效地确保大数据技术的稳定运行,充分发挥其潜在价值,确保商业决策管理环节都能最大化其价值,并推动数据资源的合理分配。

参考文献

- [1] 梁凌宇,黄文琦,袁红霞,等.基于大数据分析的非侵入式电力负荷数据分类[J].电子设计工程,2024,32(8):96-100.
- [2] 徐利棚.基于大数据分析的智能电网监控运行建设[J].中华纸业,2024,45(4):115-117.
- [3] 陈静荣.网络数据挖掘在商业决策中的应用机制探析[J].传媒论坛,2024,7(7):114-117.
- [4] 李平荣.大数据时代的数据挖掘技术与应用[J].重庆三峡学院学报,2014(3):45-47.
- [5] 刘铭,吕丹,安永灿.大数据时代下数据挖掘技术的应用[J].科技导报,2018,36(9):73-83.

Research on IDC Data Center Management Based on Intelligent Operation and Maintenance

Wenqi Wang

Guangxi Computing Center Co., Ltd., Nanning, Guangxi, 530000, China

Abstract

This paper deeply discusses the key role of intelligent operation and maintenance in IDC data center management, and conducts a forward-looking analysis of its future development. By defining the basic concepts and functions of a data center, this paper explains how intelligent operation and maintenance can improve the operation efficiency and reliability of a data center through automated tools and advanced algorithms. Intelligent operation and maintenance uses real-time monitoring, predictive maintenance and automatic troubleshooting technologies to significantly reduce labor costs and improve system stability. Looking forward to the future, with the development of artificial intelligence, machine learning and other technologies, the intelligent operation and maintenance will more accurately adapt to the dynamic environment of the data center, and further improve the intelligent level of the data center management. This paper comprehensively analyzes the management function of intelligent operation and maintenance and its potential impact on the future data center management.

Keywords

intelligent operation and maintenance; IDC data center management; data center equipment monitoring

基于智能化运维的 IDC 数据中心管理探究

王伟奇

广西计算中心有限责任公司, 中国·广西 南宁 530000

摘要

论文深入探讨智能化运维在IDC数据中心管理中的关键角色, 并对其未来发展进行前瞻性分析。通过定义数据中心的基本概念和功能, 论文阐述了智能化运维如何通过自动化工具和先进算法提高数据中心的运营效率和可靠性。智能化运维利用实时监控、预测性维护和自动化故障排除等技术, 显著降低了人力成本, 提高了系统稳定性。展望未来, 随着人工智能和机器学习等技术的发展, 智能化运维将更加精准地适应数据中心的动态环境, 进一步提升数据中心管理的智能化水平。论文全面分析了智能化运维的管理作用和其对未来数据中心管理的潜在影响。

关键词

智能化运维; IDC数据中心管理; 数据中心设备监测

1 引言

随着云计算、大数据、人工智能等新兴技术的快速发展, 数据中心作为支撑这些技术发展的基础设施变得越来越重要。传统的数据中心管理方式已经无法满足日益增长的管理需求, 智能化运维应运而生。智能化运维利用先进的技术手段, 实现数据中心设备的智能监测、故障诊断、资源优化等功能, 为数据中心管理提供了全新的视角和解决方案。

2 IDC 数据中心的概念

互联网数据中心 (IDC) 是为了满足互联网业务以及信息服务需求而构建的应用基础设施, 它是一个网络概念, 提供网络基础资源的一部分, IDC 还是一个服务概念, 为各类

企业提供服务器托管等服务。在 IDC 中, 拥有完善的设备以及完善的应用服务平台, 与互联网的连接, IDC 利用丰富的技术为客户提供互联网基础平台服务。IDC 的发展与互联网的快速发展密切相关, 随着互联网的数据量的不断增加, 对数据中心的需求也日益增长。IDC 的服务范围涵盖了各行各业, 为它们提供数据存储、处理、传输等基础设施服务。IDC 的发展推动了数字经济的发展, 也促进了各行业的信息化进程^[1]。

3 智能化运维在 IDC 数据中心管理中的管理作用

3.1 数据中心设备监测与故障诊断

智能化运维系统在 IDC 数据中心管理中通过实时监控和故障诊断功能保证设备运行的可靠性, 系统装备了多种传感器, 会不断收集网络设备的运行数据, 然后将数据实时传

【作者简介】王伟奇 (1980-), 男, 中国广西南宁人, 本科, 助理工程师, 一级建造师, 从事IDC数据中心运维研究。

输至中央监控系统，由先进的数据分析引擎进行处理。数据分析引擎使用机器学习算法分析数据，识别出设备运行中的异常模式。例如，温度异常升高或网络流量突然增大，这些往往是设备故障或性能下降的前兆。而当系统检测到潜在故障时，自动启动故障诊断程序。该程序细分为多个步骤，常见的有故障类型的初步判断、可能影响的服务范围评估、故障原因的深度分析等。故障诊断不仅依赖于实时数据，还结合设备的历史运行数据，利用模式对比来确定故障的具体性质。若系统确认是常见的故障模式，会根据预设的策略自动处理，比如调整网络配置或重启设备。在处理不明确的复杂故障时，系统会将问题标记并提醒维护团队进行手动干预。维护团队接到通知后，利用系统提供的详细诊断报告的解决方案，快速响应故障处理。而且系统会记录这一事件的所有详情，包括故障发生的时间、影响范围、处理过程和结果，这将被用于未来的系统优化^[2]。

此外，智能化运维系统还能根据设备的实际运行状态动态调整维护计划，通过预测性维护功能，系统能够预测设备潜在的故障点和可能的故障时间，从而提前进行维护，避免故障的实际发生，保障了数据中心的高效运行。

通过这一系列流程，智能化运维系统在 IDC 数据中心中扮演着关键的角色，保持设备状态的最佳化还保证了数据中心的稳定运行。

3.2 资源利用率优化

智能化运维在 IDC 数据中心管理中的资源利用率优化方面，主要通过数字孪生技术、人工智能和大数据分析等手段实现全面监控和精细化管理。数字孪生技术构建了数据中心的虚拟模型，实时反映物理设备的运行状态，通过虚拟仿真进行优化。人工智能在资源调度中发挥重要作用，基于历史数据和实时监控信息，自动调整设备运行参数提升资源利用效率。大数据分析系统整合了各类数据中心设备的运行数据，进行深度分析帮助运维人员做出科学决策。另外，通过 AI 智能巡检系统，减少了人工巡检的频率降低了运维成本，设备预测性维护系统在设备出现故障前通过数据分析进行预警，避免了突发故障带来的资源浪费。智能化管理平台实现了对数据中心的统一管理 and 全生命周期管理，提高了设备使用效率。而且，数据中心的容量管理系统通过实时监控的运行负荷情况，动态分配计算资源和存储资源，避免了资源的闲置。冷却系统的智能调控根据设备的运行状态，自动调整冷却设备的运行模式优化能效利用率。电力管理系统利用智能算法，优化电力分配减少不必要的电能消耗。

智能化运维平台集成了多种管理功能，运维人员能够实时掌握数据中心的运行状态。数据中心的智能管理系统还能够自动生成各类报表，帮助运维人员及时发现并解决潜在问题，提高数据中心的运行效率。

3.3 安全性管理

智能化运维在 IDC 数据中心管理中的安全性管理涉及

多个具体细节。一方面，实时监控是数据中心安全管理的基础，采用摄像头设备，24 小时不间断地监控数据中心的各个关键节点，能及时发现异常情况。网络安全防护同样重要，通过部署防火墙、入侵检测系统（IDS）和入侵防御系统（IPS），有效拦截外部的网络攻击。访问控制系统则通过生物识别、智能卡和密码等多重身份验证方式，严格控制人员的进出，保证只有授权人员才能进入关键区域。在数据安全方面，智能化运维系统会对数据进行加密处理并设置权限管理，使数据在传输存储过程中不被非法访问。还要定期的漏洞扫描能及时发现系统中的潜在风险并进行修补，防止安全漏洞被恶意利用。灾备系统的建立也是安全管理的重要环节，通过数据备份和异地容灾，保证在发生灾难时数据可以快速恢复业务不中断。另一方面，智能化运维中还包括对设备的健康状态监控，通过对设备的运行状态进行实时分析，提前预警可能出现的故障，避免因设备故障而引发的安全问题。应急预案的制定和演练也是不可或缺的环节，通过模拟各种突发事件，检验和完善应急响应流程，确保在实际发生紧急情况时，能够迅速地进行处置。还有一点，安全培训和意识教育也被纳入智能化运维的范畴，通过定期的培训提高运维人员的技能水平，防止因人为操作失误导致的安全事件。智能化运维系统的日志管理功能可以记录所有操作，便于事后追溯分析提高安全事件的处理效率。智能化运维在 IDC 数据中心的安全性管理中，通过多手段的安全防护措施，构建了一个全方位的安全防护体系，有效保障了数据中心的安全稳定运行^[3]。

3.4 环境监测与节能减排

智能化运维系统在 IDC 数据中心环境监测与节能减排方面实施了一套全面的管理策略，系统部署了多个传感器，持续监测数据中心的温度、湿度、空气质量和电力消耗。传感器收集的数据实时传送至中心处理单元，通过智能调控系统，数据中心的冷却和加热系统能根据实际需要自动调节，避免不必要的能源浪费。温度管理采用分区控制策略，让各个区域根据设备密度和热负荷进行精确冷却。这种分区方法允许系统对不同区域进行个性化管理，降低整体能耗。而且智能化运维系统利用预测性维护技术优化能源利用。系统通过分析设备的运行数据，预测设备可能的故障点，提前进行维护或替换，从而减少由于设备故障导致的能源浪费。另外，智能化系统还控制数据中心内的照明系统，如使用自动照明控制器在有人活动时光亮必要区域的灯光，减少电力消耗。此系统与安全监控系统相结合，能够根据人员在数据中心内的移动自动调整照明。为进一步降低能源消耗和碳排放，数据中心采用了可再生能源解决方案，如太阳能板，以及使用高效能的电源和冷却设备。智能化系统监测这些设备的运行效率，保证能源使用最大化效率。通过这种综合的策略，智能化运维系统在环境监测为数据中心提供了高效的管理解决方案，保证运行环境的持续优化与能效最大化。

例如,某大型云服务提供商部署了一套基于机器学习的故障预测系统,该系统能够分析历史实时数据,识别潜在的硬件故障。系统采用了数据挖掘技术从每日产生的数TB级日志文件中提取有价值信息,使用异常检测算法来预测故障,在测试期间,准确率高达98%显著减少了意外停机时间。对于能源管理,智能化系统通过实时数据分析优化数据中心的能源分配。采用先进的热图技术,系统对数据中心的热环境进行3D建模,分析并调整冷却系统的工作状态,以达到最佳能效比。该技术能够在保持设备最佳运行状态的同时,节省高达20%的能源消耗。网络管理方面,智能化运维通过持续的网络流量监控自动优化数据流。部署了自动化配置工具的数据中心,实现了网络设备的即时配置更新。这种自动化流程不仅提高了网络的可用性,还提高了数据传输效率。在安全监控方面,使用了深度学习算法来增强传统的威胁检测系统。能够从海量数据中学习并识别新型的安全威胁,实时更新防御策略。安全系统的部署后,未经授权的访问事件减少了40%,显著提升了整个数据中心的安全级别。

通过这些智能化运维的实施,IDC数据中心能够实现更高效、更安全、更可持续的运营。

4 展望智能化运维技术在数据中心管理中的未来发展

随着人工智能技术的不断发展,智能化运维在数据中心管理中的应用将会更加广泛。未来,智能化运维技术有望在以下几个方面取得更大突破。

4.1 自动化运维

自动化运维作为智能化数据中心管理的核心,通过自动化技术彻底改变了传统的运维模式。在实施中,数据中心部署自动化工具以实现软件和硬件的自动配置。系统自动按照预设的参数执行任务,保证部署的一致性。而且监控系统全天候不间断地收集设备运行数据,对系统性能进行实时监控,及时发现并报告性能下降或故障的迹象。这些智能监控工具可以自动诊断问题并提供可能的解决方案,甚至在多数情况下能够自主执行修复程序。随着人工智能和机器学习技术的进步,未来的自动化运维将更加智能化,能够自动调整策略以应对数据中心环境的快速变化,从而进一步提高运维效率,减少停机时间降低整体的运维成本。

4.2 预测性维护

预测性维护是指利用数据分析技术,对数据中心设备的运行状态进行实时监测,来实现对设备可能出现故障的提

前预警。未来,随着大数据的不断成熟,预测性维护将成为数据中心管理中的重要发展方向。通过对设备运行数据的深度分析,可以预测设备可能出现的故障,提前采取维护措施避免设备故障对数据中心运行的影响,从而提升数据中心的稳定性。

4.3 跨平台集成

随着数据中心技术的不断发展,跨平台集成将成为数据中心管理中的一项重要趋势。跨平台集成可以将不同平台上的数据和资源进行有效管理,实现数据中心管理的统一化。未来,数据中心管理人员可以通过跨平台集成技术,实现不同平台之间的数据共享调配,提高数据中心的整体运行管理水平。

4.4 安全防护

数据中心的安全问题一直是业界关注的焦点,随着智能化运维技术的不断发展,安全防护工作也面临着新的挑战。未来,数据中心管理将在以下几个方面加强安全防护工作:

加强边界防护:随着数据中心的规模不断扩大,边界防护工作将变得更加重要。智能化运维技术可以实现对数据中心边界的全面监控,加强对外部攻击和恶意访问的防范。

强化数据加密:数据安全一直是数据中心管理的头等大事,智能化运维技术可以通过数据加密等手段,保障数据的安全传输存储,防止数据泄露提高数据中心的整体安全性。

提升安全意识:智能化运维技术可以通过智能监控,提升数据中心管理人员对安全风险的感知能力,加强安全意识教育,使全体员工都参与到安全防护工作中来。

5 结语

智能化运维技术在IDC数据中心管理中发挥着越来越重要的作用,涉及设备监测、资源利用率优化、安全性管理、环境监测与节能减排等方面。未来,随着新兴技术的不断发展,智能化运维技术将会迎来更加广阔的发展空间,为数据中心的安全稳定运行提供更加有力的支持。

参考文献

- [1] 王清江.数据中心设备及机房的智能化运维管理分析[J].自动化应用,2022(7):157-159.
- [2] 苏剑.数据中心设备及机房的智能化运维管理[J].中国管理信息化,2021,24(22):97-98.
- [3] 涂祥宇,熊慧亮.浅谈数据中心设备及机房的智能化运维管理[J].中国新通信,2020,22(16):70.

Network Public Opinion Analysis System Based on Big Data Technology

Liying Wang

Guangzhou College of Technology and Business, Guangzhou, Guangdong, 510850, China

Abstract

With the wide application of China's Internet, the rapid development of mobile phone, wireless Internet access and other technologies, the modern Internet has become an important means of human information communication and cultural communication. The network is increasingly infiltrating into our daily life. At the same time, the daily data generated on the Internet grows in geometric times, making it more difficult to analyze public opinion, thus entering the era of big data and increasing the difficulty of public opinion monitoring. Moreover, online public opinion work is the most difficult work in the world, especially in China. If it is not strictly regulated, it may endanger social stability. This paper, first of all, analyzes the current domestic public opinion, and defines the concept and characteristics of big data. It plans to organically integrate public opinion analysis and big data technology to build a set of public opinion analysis system under the big data environment.

Keywords

big data; network public opinion; analysis system; design

基于大数据技术网络舆情分析系统

王丽颖

广州工商学院, 中国·广东广州 510850

摘要

随着中国互联网的广泛应用, 移动电话、无线上网等技术的飞速发展, 现代互联网已经成为人类进行信息沟通、文化传播的重要手段。网络正日益渗透到我们的日常生活中, 同时, 互联网上每日生成的数据以几何倍数增长, 使得对舆情进行分析变得更加困难, 从而进入了大数据时代, 也加大了舆情监测的难度。而且网络舆情工作是全球最困难的工作, 特别是中国, 如不严加管制, 则可能危及社会安定。论文首先对当前国内舆情进行了分析, 并对大数据的概念、特征进行了界定, 拟将舆情分析和大数据技术有机地融合起来, 构建一套大数据环境下的舆情分析体系。

关键词

大数据; 网络舆情; 分析体系; 设计

1 大数据技术和网络舆情的概述

1.1 大数据技术综述

大数据技术的本质就是, 数据的数量和规模都比以往要多得多, 并且不能通过现有的一些软件来对这些数据进行整理和分析。通过运用大数据技术, 政府部门和企业可以有效地对大数据进行分析, 并由此制定出一种有效的商业开发规划与管理方式, 大数据的规模大, 类型多, 价值低, 密度低, 流通速度快。在其他国家, Hadoop 系统已经具备了专门的

数据处理技术所无法比拟的可靠、高效和可扩充性, 这个平台包含了很多部件, 比如多个存储节点, 它们能够从一个节点中采集并处理来自各个节点的数据, 很多高性价比的电脑都能根据系统的需求增加一个处理器节点。

1.2 网络舆情综述

网络舆情一般是互联网用户在面对社会热点新闻时, 尤其是涉及到自身利益和国家政策等方面的情感和观点的一种反映。这类舆情带有某种偏向性, 舆情是各种意见、态度的总和, 其特征是: 具有很强的广泛性, 舆情能够在全国范围内很快地扩散开来, 参与人的种类也很广; 突发事件的特点: 在一个区域内, 一旦有突发事件发生, 就能作为舆情的源头; 主观方面表现为: 舆情内容、舆情意见主观性强; 多元化的特点表现为: 参与人对信息的看法、观点的不同, 以及对信息的传播、表达等方面的不同。随着时代的进步, 舆情的数量与内容也逐年增多, 如何及早识别并妥善处理舆

【基金项目】广州工商学院2022年校级科研项目(重点项目)《基于Single-Pass算法增量式聚类的网络舆情挖掘与分析研究》(项目编号: KYYB202228)。

【作者简介】王丽颖(1991-), 女, 硕士, 讲师, 从事人工智能、数据挖掘研究。

情,已成为当前政府部门必须面对的重大问题。

2 大数据技术下网络舆情分析系统的设计原则

在大数据的背景下,对舆情分析系统的要求如下。

2.1 信息抓取必须全面

当前,互联网舆情具有涉及面广、种类繁多、网页数目多等特征。所以,舆情的收集需要有能力收集来自各种结构化和半结构化的舆情,建立一种新的储存和网络爬虫器^[1]。

2.2 确保有效的分析

优秀的网络舆情分析软件能够帮助政府部门、企业对网络舆情中的热门主题、热点事件进行及时的追踪,从而对网络舆情作出及时的应对。当出现消极意见时,能够快速地采取恰当的措施来恢复和维持稳定。

2.3 结构设计

①采集模式。并通过网络爬虫技术实现对互联网上的站点的快速、大范围的抓取。系统会为建立表格和黑名单而指定 URL 收集点。基于特定需求,通过搜索、储存“白名单”和“黑名单”站点,可以有效地提高搜索效率,避免不必要的操作。

②信息预处理模式。它的工作是对采集来的原始页面进行转化,再将经拷贝、去噪后得到的格式文本存入数据库,抽取文字信息,对其进行分段,采用中文切分方法产生单项。在此基础上,利用矢量空间、概率等方法,对网络舆情进行特征抽取,形成面向舆情的文本矢量集合。

③舆情分析模式。这是舆情分析中的大数据,其主要作用是实现话题识别,文本趋势分析,热点挖掘等。该方法的基本思想是:利用话题识别技术对文本矢量进行解析、学习、聚集与相同事件有关的海量文本信息,从而发现主题,利用话题追踪方法,对向量化的文本进行检索,并通过相似度来判断话题的相关性,从而实现对新文本的分类。在此基础上,利用微机计算技术对文本中的观点、态度、情感及非事实等信息进行分析,获取恰当的文本语义,为监管部门和政府部门及时发现不良舆情提供依据。通过建立话题的统计,也能识别出一些重要的信息,例如:来源、评论的数量、演讲的时间。

④公众意见报告模式。负责向有关政府部门及监管部门发布调查结果,为其制定政策提供依据。

3 大数据技术下的网络舆情分析系统

3.1 系统功能框架

构建以大数据为基础的网络舆情分析系统。该系统采用分布式文件系统(Distributed File System, HDFS)对数据进行存储与管理。MapReduce 程序设计模型把大型的工作分割成多个小型的工作,这些工作在多台服务器上平行地运行。最后,利用分布式计算框架,归纳各子任务的效果,并进行情感评价与倾向性分析。

3.2 系统功能

3.2.1 数据收集功能

最基本的一个功能就是搜集各种网络上的舆情消息,比如微信、微博等等。在大数据背景下,在利用常规搜索引擎爬虫确保数据完整的基础上,通过有针对性的爬虫来提高信息收集的效率和准确性,还可以使用黑白名单等手段,将合法的链接保留下来,从而更好地实现新的搜索。在对网站进行信息抓取的过程中,首要的工作就是对网站中的文本内容和栏目列表进行采集,在搜集资料的过程中,为下一步的工作打下了良好的基础。

3.2.2 数据处理模块

互联网具有数据量大、时效快和信息多样化等特征,网页的信息结构与内容多种多样,包含文字资料、图片资料、声音资料及影像资料。由于所收集的来源数据无法达到直接的解析需求,因此必须对所收集的数据进行预处理与向量化,在互联网舆情的研究中,必须先对互联网舆情进行预处理,然后再进行向量化,对原始文本行集进行去重、去噪,再根据文字的特性选择对原文本集进行量化,从而获得一组文本集。由于所收集到的资料在结构上存在着不同的差别,因此有必要对这些不相干的界面资料进行整理,并且保存界面标题、内容摘要和发布时间等关键信息,在此过程中,为保证数据完整性,需要对遗漏的数据进行删补,论文拟采用 MapReduce、分词工具等方法,对已有格式的文本进行并行处理,抽取具有代表性的词汇特征,并将其进行聚类,生成相应的文本矢量,并将其存入分布式 HDFS 文档构件中。

3.2.3 舆情数据采集技术

舆情信息收集的方法,就是确定舆情的议题,确定舆情的出发点,舆情数据采集是网络舆情分析中非常重要的一步,可以为以后的数据处理和数据分析提供依据。舆情数据收集技术可以从最初的 URL 中,获得里面的网页信息,之后,把这些网页信息保存在一个本地的系统里,并且通过对该网页的结构和内容的解析,提取这个网页的连接,然后将其转换为一个新的 URL。目前,主流的爬虫技术主要有:主题爬虫和增量爬虫等。每种爬行方式各有其特点与优势,选择爬行方式时应结合网络民意的实际情况^[2]。

3.2.4 舆情数据预处理技术

在使用网页爬行器来获取页面时,其实际的结构和实际的内容存在较大差异,导致大量的数据信息无法满足用户的需要。在这种情况下,为了确保对网络舆情进行有效的分析,必须对其进行预处理。舆情资料的预处理就是要对互联网上的舆情做好准备,排除噪音、重复等问题。例如,在词汇预处理方面,应该充分利用中文分词技术,对采集到的文字进行高效的切分,然后把它转化成包含各种词的集合,能有效地剔除语句中的停顿成分,并对语篇中不同词语的使用情况进行统计,使文本的预处理工作变得更加容易。

3.2.5 舆情智能分析技术

舆情智能化研究是当前互联网舆情研究领域的一个重

要研究方向,其研究内容包括:主题的识别和追踪,热点主题的发现,以及文本倾向性的挖掘。主题辨识则是运用预先设定好的数据向量集,再利用计算机自动完成主题识别。在一个文件中对同类事件进行总结,并确定它们之间的舆情主题,通过 Hadoop 对文本数据进行分块,并将其与核心文件一同传送到 map 函数完成相关运算。Map 可以将小块内的数据分布给离它很近的中间点,然后用键值对的方法,把它传送给 Reduce,这样就可以对计划进行平均化,得到一个新的聚类中心。主题追踪则是基于最近一次更新的向量文本的迅速查找,并通过相似度的计算来实现,并评价其与当前主题的相似程度,如果相似度满足要求,此类文字就会被归类为主题,如果不符合要求,它就会被分类成一个完全不同的主题。在此过程中,要对主题的评论数、转发量和赞数进行分析,再计算受欢迎指标,并与受欢迎指数进行综合,对热点主题进行筛选^[3]。

3.2.6 舆情预测预警技术

舆情预测与预警技术是对舆情进行智能化分析的结果。在对热点新闻、事件或舆情进行监控时,一旦用户到达设置的告警数值,将自动产生相关舆情报告,并以电邮或短信方式向工作人员发送,得到通知,员工就能高效的解决这一问题。

3.2.7 报告应用模块

报告应用模块以对舆情信息资料进行推送,并对其进行分析,以文本、图表、报告的形式,实现舆情综述、舆情风险监控、热点事件分析、大风险预警可视化等功能。其中,舆情概略完成了用户监控、本地相关信息、本地负面信息、预警信息推送以及本地媒体量的统计展示。舆情风险监控按照主题的热门程度将其划分为不同的类别,并将事件的标题、原始链接、来源、已发布事件、热度值以及发展趋势等信息显示出来,并将其与历史上相同的重要事件或者国家范围内的相似事件进行对比和展示,并通过对其热度与趋势的分析,识别舆情风险。

3.3 系统的设计

在特定模型中,应该考虑如下问题:

①信息的采集必须保证完整。目前,网络舆情呈现出规模大,类型多,网站多的特征。所以,在实际的建模过程中,要充分地收集数据是非常必要的。确保网页、声音、图片、结构化或半结构化的资料,以及最佳储存及网络爬行器优化。

②确保快速响应公众舆情。今天,随着网络的迅猛发展,在互联网上,舆情呈现出了一种灵活多变的特点,而常规的舆情分析模式,已不再适用于目前的情况。特别是在现今的发展环境下,网页与图像中所隐含的信息难以被及时地发现与捕捉,进而,在建模时,需要充分考虑用户对舆情的响应能力,并对隐含的信息进行有效搜集,从而达到准确的网络

舆情引导。

③保证分析结果的正确性。一个高效的网络舆情分析软件能够帮助企业及主管部门迅速了解到网络上的热门主题及相关信息,通过这种方式,既能在舆情爆发前得到有效的遏制,又能使舆情的影响力得到最大限度的发挥。因此,在进行模式设计时,必须对舆情信息进行充分了解,才能保证模式的合理性。

4 基于大数据的互联网舆情分析体系的不足之处

在一个具体的应用中,下面三条必须被遵循:

①资料完备性。大数据环境下,互联网舆情具有规模巨大、数量庞大、内容多样等特征,特别是文本、图像、音频、视频等多种应用场景下,因种种因素而未能被及时发现和捕获的隐性信息,为此,如何根据实际情况,不断地创新和优化存储和网络爬虫技术,建立科学合理的数据抓取方法,在保证数据安全的前提下,高效地获得隐藏信息。

②警告的准确性。针对现有的基于数据处理生成的预警系统,特别是在网络舆情的传递中,有许多人工合成的谣言,很难分辨真伪,这时候就必须进行人工干预和判断。

③合理的解析,一个高效、智能化的网络舆情分析工具,能够帮助用户在网络舆情爆发前,高效地对网络上纷繁复杂的信息进行分析,从而发现热点、追踪发展趋势,实现对舆情的科学、有效的管控。所以,在对其进行分析时,要特别注意算法的选取^[4]。

5 结语

大数据本身的特点,使得大数据具有很强的应用价值,大数据能够提供更加精确和全面的信息,能够挖掘出之前没有被挖掘出来的信息的规律与联系,从而帮助使用者快速作出决定。以大数据为基础的互联网舆情分析系统,在对舆情进行预警、预测和报告等方面有着独到的特色,体现出大数据的独特优势。它能有效地采集、处理、分析、展示大量的文本、图像、音频、视频等数据,对互联网舆情的处理起到了至关重要的作用。当然,在具体的实施过程中,必须对数据的准确性、及时性、完整性等方面进行全面的考虑与优化。

参考文献

- [1] 代青松,李泽华.基于大数据技术网络舆情分析系统[J].电脑编程技巧与维护,2024(3):72-75.
- [2] 陈怡菲,李雨静,肖金兰,等.基于大数据技术网络舆情分析系统[J].网络安全技术与应用,2023(9):62-63.
- [3] 张晓飞.基于大数据技术的网络舆情分析系统研究[J].无线互联科技,2021,18(2):17-18.
- [4] 孙昊.大数据技术下的网络舆情分析系统研究[J].自动化与仪器仪表,2018(8):26-28.

Research on Network Fault Prediction and Diagnosis Technology Based on Artificial Intelligence

Chen Sang

Tianfu New Area General Aviation Vocational College, Chengdu, Sichuan, 610000, China

Abstract

With the rapid development of the Internet, network fault diagnosis and prediction have become a necessary step to ensure the stability of network operation. However, due to the complexity of the network, the traditional fault detection methods often cannot meet the requirements of real-time fault diagnosis. Therefore, this study uses the AI technology to study the prediction and diagnosis of network faults. First, a large amount of network operation data is collected and analyzed, and the deep learning method is used to preprocess the data. Then, models, such as deep neural network (DNN), convolutional neural network (CNN) and long-and short-term memory network (LSTM), are used for fault prediction and diagnosis. The relevant experimental results show that, by combining various network parameter data, the artificial intelligence model has a high prediction accuracy for the future network faults, and the speed of fault identification and positioning is significantly higher than that of the traditional methods.

Keywords

AI; network fault prediction; fault diagnosis; deep learning; stable network operation

基于人工智能的网络故障预测与诊断技术研究

桑晨

天府新区通用航空职业学院, 中国·四川成都 610000

摘要

随着互联网的快速发展,网络故障诊断和预测成为确保网络运行稳定的必要步骤。然而,由于网络的复杂性,传统的故障检测方法往往不能满足实时故障诊断的要求。因此,本研究利用人工智能技术进行网络故障的预测与诊断研究。首先,搜集并分析大量网络运行数据,采用深度学习方法对数据进行预处理。然后,运用模型,如深度神经网络(DNN)、卷积神经网络(CNN)和长短期记忆网络(LSTM)等人工智能技术进行故障预测与诊断。通过相关实验结果表明,通过结合各种网络参数数据,人工智能模型对未来网络故障的预测准确率较高,对故障的识别和定位速度比传统方法明显提升。

关键词

人工智能; 网络故障预测; 故障诊断; 深度学习; 网络运行稳定

1 引言

互联网早已成为我们生活中不可或缺的一部分,有效稳定的网络运行是保证日常生活和企业工作顺畅进行的必要之处。然而,由于网络的复杂性,确保其稳定运行一直以来都是个难题。一旦发生故障,可能会引发各种各样的问题,如跨国公司的交流困难、人们在线支付的阻碍,甚至影响到网络医疗系统的运行,由此可见,网络运行稳定性的重要性。过去,我们通常依赖于传统的故障检测方法来进行网络故障的诊断和预防,然而,由于其固有的局限性,这些方法往往不能满足我们对实时故障诊断的需求。近年来,随着人工智能技术的发展,我们发现了新的解决策略。本研究旨在探讨如何利用人工智能进行网络故障的预测与诊断,并对其可能

的应用进行深入讨论。

2 网络故障的形态与影响

2.1 网络故障的类型和特点

网络故障的类型和特点广泛且复杂^[1]。类型上,网络故障可以划分为设备故障、链路故障、配置故障和安全故障等多种。其中,设备故障通常是由硬件问题引发,链路故障则常常由于物理连接或协议错误引起。配置故障和安全故障是因网络设备设置或网络安全措施不当导致。

这些故障的特点也各异。例如,设备故障与链路故障往往突发且难以预知,可能导致网络中断或严重延迟。配置故障和安全故障则可能对网络运行产生慢性影响,直到问题被发现并修复。而且,这些网络故障往往相互影响,一个故障往往会引发其他类型的故障,增加了网络故障诊断的难度。了解网络故障的类型和特点,有助于更准确地识别和处理网络故障,保证网络的稳定运行。

【作者简介】桑晨(1996-),男,中国四川成都人,本科,助理工程师,从事计算机网络技术研究。

2.2 网络故障的常见产生原因

网络故障的常见产生原因可以归为以下几点：一是网络设备故障。这包括服务器、交换机、路由器等网络硬件设备的故障、老化、过热、电源问题等。二是网络配置错误。由于网络管理人员的操作失误或知识欠缺，导致网络配置有误，影响网络的正常工作。三是软件故障或病毒攻击。网络中运行的软件存在故障，或者被恶意病毒、木马、勒索软件等攻击，会对网络数据进行损害，从而影响网络的运行。四是网络负载过大。当网络的需求超过了网络的最大处理能力，就会出现网络拥堵，从而导致网络故障^[2]。五是网络运营商出现故障。如ISP线路故障等。除此之外，自然灾害等不可抗力因素也可能导致网络出现故障。以上多种因素并存，导致网络故障异常复杂，恰当有效地故障诊断成为网络运维的重要课题。

3 基于人工智能的网络故障预测与诊断技术

3.1 采集并分析网络运行数据的方法

采集并分析网络运行数据是基于人工智能的网络故障预测和诊断的重要步骤。通过部署高级监控工具在网络设备上收集各类数据，包括网络流量、速度、延迟等，也需要收集设备的硬件和软件状态数据，如内存使用情况、CPU占用率、设备温度等。采集的数据越全面，预测和诊断的准确率越高。对收集的数据进行预处理，包括数据清洗和标准化，以便进行后续的分析。利用统计和深度学习技术对数据进行归纳和总结，得出网络运行的模式和规律。其中，可以采用时间序列分析、趋势分析等方法提取出预测网络故障的关键特征。以上过程涵盖了采集并分析网络运行数据的主要方法，为后续基于人工智能的网络故障预测与诊断提供了重要的数据基础^[3]。

3.2 深度学习在网络数据预处理中的应用

深度学习在网络数据预处理中发挥了重要的作用。为确保数据的可靠性和质量，必须对原始的网络运行数据进行有效的预处理。深度学习算法通过自我学习和自我调整的方式，能够有效地识别和提取数据特征，进而提高预处理效率。具体来说，深度学习可以通过自编码器方式进行特征学习，该方法可以从未标记的数据中学习出有用的特征，从而减少对手动特征设计的依赖，大幅提升了预处理效率。深度学习的稀疏编码技术，可以消除数据中的冗余信息，提高数据有效性。这对于网络故障预测与诊断研究来说至关重要，因为这可以帮助模型更好地理解网络运行数据的模式，进而实现对网络故障的准确预测和快速诊断。

3.3 深度神经网络、卷积神经网络和长短期记忆网络在故障预测与诊断中的应用

深度神经网络（DNN），卷积神经网络（CNN）和长短期记忆网络（LSTM）等人工智能模型是网络故障预测与诊断技术的主要工具。演变出多层隐藏层的DNN，通过

自我学习，能够准确预测网络故障。CNN的特点在于通过卷积运算实现模式匹配，能有效输出参数对关键特征进行捕捉，从而识别出网络故障的表现形态，达成诊断目标。LSTM则是一种时间序列分析方法，其优点在于能分析网络长周期内的故障模式，优化故障预测。这三种模型相结合，可以更全面地进行网络故障的预测与诊断，从而显著提高网络维护效率，提供强有力的技术支持。

4 实验分析与评估

4.1 实验方法与实验环境介绍

实验方法主要为网络运行数据的采集和分析，以及人工智能模型的训练和测试。使用数据采集工具获取大规模的网络运行数据，通过深度学习方法对这些数据进行预处理，以便模型更好地学习和理解。预处理的步骤主要包括数据清洗、缺失值处理、数据缩放等。

选择深度神经网络（DNN）、卷积神经网络（CNN）和长短期记忆网络（LSTM）等模型，用预处理后的数据进行训练。具体的训练过程将依据模型的特点和训练数据的性质，选择合适的优化算法和损失函数，以得到具有良好泛化能力的模型。

实验环境为一台配置有高性能处理器和高容量内存的计算机，上述所有实验均在此环境下完成。使用的工具库包括深度学习框架TensorFlow，数据处理库numpy和pandas等。这样的环境确保了实验的高效进行和结果的准确性。

4.2 实验结果分析

对比使用深度神经网络（DNN）、卷积神经网络（CNN）和长短期记忆网络（LSTM）进行网络故障预测与诊断的实验结果，均取得了显著的成果。深度神经网络对复杂网络故障的识别与定位速度提升明显，其预测准确率较高，特别是在辨别复杂型网络故障上，展现了超乎寻常的优势。卷积神经网络在处理大量网络参数数据时表现出了极高的运行效率，对网络故障的预测准确率也有显著提升。而长短期记忆网络的复杂度相对较低，实际应用中更为灵活。整体来看，人工智能技术在网络故障检测与诊断上已经明显超越了传统手段。

5 未来研究方向

5.1 高度复杂网络环境下故障预测与诊断的研究

在未来网络的调查与研究领域，高度复杂的网络环境下的故障预测与诊断显示出极其重要的实用价值。当前虽然已有的人工智能模型能在一定范围内实现网络故障的有效预测与诊断，但面对维度更高、结构更复杂的网络环境，这些模型可能力不从心。未来的研究将聚焦于为高度复杂网络环境中的故障预测与诊断提供解决方案。研究将尝试优化和改良现有的人工智能模型，提高其泛化能力，使其能更好地处理复杂网络故障的预测与诊断；也将寻求新的人工智能技术和方法，求新求异，力图使网络故障预测与诊断技术取得

新的突破。

5.2 人工智能技术在网络故障预测与诊断的深层次应用

人工智能技术在网络故障预测与诊断的深层次应用是未来研究的重要方向。尽管目前的系统已经能通过人工智能模型对网络故障进行较为准确的预测与诊断，但仍面临许多的挑战和待解决的问题。如何通过更深入的模型学习，按照网络故障的实际产生过程来进行预测和诊断，将更接近真实的网络环境。以往的研究工作许多停留在表面层次，对网络故障的预测和诊断体验尚未深化到网络操作层面。未来的研究将尝试通过模型的深度学习，对网络操作过程进行模拟，实现对故障发生前的预测和故障发生后的原因诊断。这既需要大量的网络操作数据，又需要高效地学习算法。如何实现高效的数据收集和优化算法，将是未来研究中的一个重要内容。

面向未来，将进一步地深化这一领域的研究与应用，剖析和利用人工智能技术在异常检测和自我修复系统中的优势，有望实现网络运行的高效稳定，提升网络故障处理的准确率和效率。

6 结论

6.1 人工智能在网络故障预测与诊断中的应用明显改善网络故障处理的效率

人工智能在网络故障预测与诊断中的应用已经取得了显著的效果。在实际操作过程中，人工智能模型通过驱动和优化现有的网络运行数据，使得故障预测和诊断的准确度得到提高，从而直接提升了网络故障处理的效率。实验证明，相较于传统方法，这种新型技术大幅度降低了 the identification and diagnosis time of network failures 的时间成本，能够将繁复的网络故障处理过程简化，提高执行效率。而且，这种技术由于其自我学习和自我调整的特性，使得预测和诊断更倾向于准确，从而减少了后期维修的工作量，大大提高了工作效率。总的来说，人工智能的应用不仅具有较高的预测准确率，而且在网络故障处理的速度和质量方面都有显著的提升，展现了强大的实用价值。

6.2 人工智能技术为未来网络诊断设计提供新方向

人工智能技术对网络故障预测与诊断体系具有深远影响。其能自动学习并精准预测网络故障，明显提高了故障处理的效率。由深度神经网络（DNN）、卷积神经网络（CNN）和长短期记忆网络（LSTM）等模型组成的预测诊断系统，

以高精度和高效率优化故障诊断流程，为网络维护人员在识别、定位故障时提供强大的技术支持。未来，人工智能还将与更多领域结合，包括异常检测、自我修复和安全防护等，为网络故障预测与诊断技术提供全新发展方向。基于人工智能的网络故障预测与诊断，不仅解决了现有网络环境中出现的各种问题，更引领了未来网络诊断设计的新篇章，有望推动整体网络环境更加稳定、高效地运行。

6.3 基于人工智能的网络故障预测与诊断体系的建立对构建稳定的网络环境的重要性

构建基于人工智能的网络故障预测与诊断体系，对于构建稳定的网络环境具有至关重要的意义。一方面，人工智能技术可以实现对大量复杂、多维度的网络运行数据的快速分析和处理，有效提升预测与诊断的速度和准确性，相比传统的故障处理方式，具有明显优势。基于人工智能的模型具有出色的泛化性能和适应性，能够应对各种复杂情况，提升系统的鲁棒性。再者，人工智能模型具有较低的计算复杂度和较高的运行效率，适合大规模的实际应用，有利于提高网络环境的稳定性。借助人工智能技术，一方面可以减少故障发生的频率和影响程度，另一方面，能够快速准确地定位并处理故障，提高网络系统的稳定性和可靠性，构建出健壮、高效的网络环境，这对于满足日益增长的网络需求，保障网络安全稳定运行，无疑具有重大的实用价值和广泛的应用前景。

7 结语

论文的研究主要基于人工智能技术，预测和检查网络是否会出问题。这个研究发现，人工智能在这个方面的表现非常好，能准确地发现未来可能出现的问题，有助于我们提前解决。但是这个方法也有一些不足，如在数据较少或很难得到数据的情况下可能效果不好。未来的研究将继续寻找解决这个问题方法，同时还要研究如何针对不同的网络环境设计和实施人工智能检查系统。这篇研究告诉我们，人工智能在预测和解决网络问题方面有很大的潜力，有机会让网络运行得更顺畅。

参考文献

- [1] 曾毅. 人工智能技术在网络故障诊断中的运用[J]. 信息与电脑, 2021, 33(21).
- [2] 金少斌, 陈曦, 金超标. 基于人工智能的5G网络故障诊断方法研究[J]. 中文科技期刊数据库(文摘版)工程技术, 2023(3).
- [3] 孙剑斐. 浅谈网络故障诊断技术中人工智能技术的应用[J]. 电脑知识与技术: 学术版, 2019, 15(7X).

AI Practice from the Perspective of the Era of Big Data

Chun Zhang

Tianjin Binhai Vocational College of Automotive Engineering, Tianjin, 300352, China

Abstract

In this era of big data, the research of artificial intelligence has gradually begun to emerge. How to effectively use these huge data to promote and popularize artificial intelligence technology has become a focus of research with the attention of The Times. In this paper, we will make some simple exploration of the progress of artificial intelligence, the focus of research, the superiority of its use, the operating environment and its future possibilities in this big data environment. For the research of artificial intelligence in the era of big data, in order to continuously optimize its application services, and realize its efficient integration with big data.

Keywords

the era of big data; AI; application practice

基于大数据时代视角下人工智能实践

张纯

天津滨海汽车工程职业学院, 中国·天津 300352

摘要

在这个大数据的时代, 人工智能的研究已经渐渐开始崭露头角。如何有效地使用这些庞大的数据来推广和普及人工智能的技术, 已经随着时代的关注, 变成了一个研究的焦点。论文对大数据环境中, 人工智能的进步情况、研究的焦点、其使用的优越性、运作的环境以及其未来的可能性进行一些简单的探索。对于在大数据时代中的人工智能相关领域的研究, 以便持续优化其应用服务, 并且实现其与大数据的高效融合。

关键词

大数据时代; 人工智能; 应用实践

1 引言

在人工智能领域的研究取得了显著的进步, 并且其应用已经深深地渗透到我们的日常生活中。在这个数据爆炸的时代, 我们需要寻找一种方法来最好地运用这些庞大的数据, 从而使我们的生活和工作更加方便。因此, 将这些数据和大数据技术紧密联系起来, 充分利用其高速、精确的计算特性, 是一个非常好的方式来解决现实中的问题, 从而推动社会的进步。

2 大数据时代人工智能实践的兴起

2.1 大数据时代人工智能实践的背景

人工智能技术得到了广泛应用, 在大数据时代背景下, 人工智能的应用场景越来越多样化, 从智能家居到自动驾驶汽车再到医疗保健领域都有着广泛的发展和应用, 而这些应用背后离不开大数据的支持与推动, 大数据时代是信息化社

会发展的新阶段。它指的是以数据为核心的新型经济形态, 通过大规模的数据采集、存储、分析和处理实现价值的最大化, 在这个过程中, 人工智能技术发挥了至关重要的作用, 大数据时代人工智能实践的出现也催生了一系列新的商业模式, 如智慧城市、智慧交通、智慧物流, 这些新型商业模式不仅改变了人们的生活方式, 还为企业提供了更多的商业机会和发展空间^[1]。

2.2 大数据时代人工智能实践的特征

大数据时代的到来, 在大数据时代人工智能实践中, 出现了一些新的特点和趋势, 数据量越来越大, 这使得机器学习算法更加复杂化, 需要更多的计算能力和存储空间; 数据来源多样且分布不均, 因此需要采用更复杂的模型进行处理, 伴随云计算的发展, 人工智能应用场景也在不断扩大, 例如智能客服、自动驾驶等。大数据时代人工智能实践的特点是多样化、复杂性和规模性, 只有通过不断地创新和发展, 才能更好地应对这些挑战并取得更好的成果。

2.3 大数据时代人工智能实践的意义

在大数据时代背景下, 人工智能不仅可以帮助企业更好地管理和分析数据, 还可以为社会提供更加智能化的服务, 人工智能实践的重要性日益凸显。从商业角度来看, 人

【作者简介】张纯(1988-), 女, 满族, 中国河北承德人, 硕士, 助教, 从事大数据、人工智能等在软件工程领域的应用研究。

工智能能够提高企业的效率和竞争力，通过对大量数据进行深度挖掘和分析，企业可以更准确地了解市场需求和客户偏好，从而制定更有针对性的产品或服务方案。人工智能还能够自动化一些烦琐的工作流程，减少人力成本并提升生产率。也可以为人类带来更多的便利。例如，智能家居系统可以通过语音识别和机器学习实现自动控制家电设备的功能；医疗领域中，AI技术可以用于疾病诊断和治疗等方面，它将深刻影响我们的生活和社会发展。

3 大数据时代人工智能实践的主要内容

3.1 大数据时代人工智能实践的主体

在大数据时代的人工智能实践中主体是数据，从传统的文本数据到图像识别、语音识别以及自然语言处理等方面的数据，都需要大量的数据支持才能实现有效的AI模型训练和预测。数据采集和处理在大数据时代的背景下，数据的获取与处理变得更加重要。通过对大量数据进行收集、清洗、分析以及建模等操作，可以获得更准确的数据结果。机器学习算法随着技术的发展，机器学习算法越来越多样化，包括监督式学习、无监督式学习、半监督式学习等等。这些算法可以用于各种应用场景，如图像识别、自然语言处理、语音识别。深度学习是近年来发展最快的技术，它能够模拟人脑神经网络结构并实现高级抽象特征提取。智能决策系统：利用人工智能技术构建出一套智能化的决策支持系统，以帮助人们做出更好的决策^[2]。云计算大数据时代需要借助云计算平台的支持，以便更好地存储、管理和共享海量数据资源。安全保障由于大数据时代中涉及大量的敏感信息，因此必须采取相应的安全措施保护用户隐私。

3.2 大数据时代人工智能实践的客体

随着技术的发展和应用场景的变化，人工智能也逐渐从单纯的理论研究走向了实际应用领域，我们将探讨大数据时代的人工智能实践的主要内容以及其对人类社会的影响。数据源包括各种类型的数据，如文本、图像和声音。这些数据可以来自互联网、社交媒体和其他来源。算法模型用于处理和分析数据的数学模型。硬件设备计算机、存储器、网络等硬件设备是实现人工智能应用的关键组成部分。软件平台提供开发工具，如机器学习框架和自然语言处理库。人员资源需要有专业的技术人才来进行数据收集、清洗、建模和部署。政策法规：在实施人工智能项目时，必须遵守相关的法律法规和道德规范^[3]。商业模式企业要考虑如何将人工智能应用于商业领域中，并制定相应的商业模式，社会影响人工智能的应用可能对社会的各个方面产生深远的影响，比如就业市场、隐私保护等方面。

3.3 大数据时代人工智能实践的载体

在大数据时代的人工智能实践中，载体是至关重要的，它可以提供一个平台来收集和海量数据，并为机器学习模型进行训练和优化，云计算、物联网以及深度学习技术都

是大数据时代人工智能实践的重要载体，云端计算提供了大规模的数据存储与分析能力；物联网则通过连接各种设备和传感器实现对环境的实时监测；而深度学习则是一种高效的人工智能算法，能够从大量数据中学习特征和模式，这些载体不仅提高了人工智能应用的质量，也推动了其发展速度。

4 大数据时代人工智能实践的问题和应对措施

4.1 大数据时代人工智能实践存在的问题

计算机网络系统中的网络控制与监控功能还存在着诸多的问题，为了保证计算机能够准确、及时地获取相关的信息，及时、有效地解决突发的问题、制定出相应的对策，数据质量由于大量数据的来源多样，数据的质量难以保证。这导致了机器学习模型的训练效果不佳。数据隐私保护在处理大规模的数据时，需要对个人隐私进行严格保护，以避免泄漏敏感信息。同时，也需要注意数据安全问题。算法透明度对于一些复杂的机器学习算法，很难理解其背后原理，缺乏可解释性。技术门槛高对于非专业人士来说，使用大型机器学习框架可能存在困难，应用场景限制目前，人工智能的应用领域还比较局限，无法广泛推广到各个行业中。

4.2 大数据时代人工智能实践的应对措施

为了保证在计算机网络系统中的网络控制与监控功能均能够正常发挥其作用，我们必须确保电脑能够精确且迅速地收集相关的数据，并能迅速且高效地处理紧急情况，并提供适当的解决方案。数据安全措施在处理大量数据时，需要采取严格的数据保护措施来防止数据泄露。这包括加密技术、访问控制、身份验证以及其他安全措施。算法优化为了提高机器学习模型的表现能力，必须对算法进行不断优化，以适应不同的场景和应用。多源数据融合利用多个来源的数据集可以帮助AI系统更好地理解复杂的问题，并提供更准确的结果。深度学习是当前最流行的人工智能技术，它能够从大量的数据中提取出有用的信息。云计算平台使用云端资源可以大大降低成本，同时为企业提供了更好的计算性能和更高的可靠性。可扩展性要保证系统的可扩展性和灵活性，因此应该选择一个具有良好扩展性的解决方案。跨领域合作通过与不同领域的专家合作，我们可以获得更多的知识和技能，从而推动人工智能的发展。

4.3 大数据时代人工智能实践的保障机制

在大数据时代的人工智能实践中，保障机制是至关重要的，为了确保数据安全、隐私保护以及算法公正性等方面，需要采取一系列有效的措施来保证人工智能系统的正常运行，应该建立完善的数据库管理制度，以防止未经授权的人员访问敏感数据；数据隐私保护在大数据时代，由于大量的个人信息被收集到数据库中，因此需要加强对数据隐私的保护；算法公正性为了保证机器学习模型的公平性和准确性，必须建立一套完整的评估体系来检测模型的偏差；数据安全要确保数据存储和传输过程中的数据安全性，防止黑客攻击

或恶意篡改,数据质量控制要定期检查和更新数据的质量,以避免出现错误导致的不良后果。

5 大数据时代人工智能实践的价值

5.1 人工智能实践的经济价值

随着大数据时代的到来,人工智能在各个领域得到了广泛的应用。从经济角度来看,人工智能的实施可以带来巨大的效益,人工智能能够提高生产效率和降低成本。通过自动化生产线、智能物流系统等技术手段,企业可以在更少的人力资源的情况下实现更高的产出水平,还可以为企业的决策提供支持,使企业能够更好地应对市场变化并做出正确的决策,人工智能还能够为企业带来更多的创新机会,促进其业务发展和持续增长,我们可以看到人工智能实践对企业的经济价值具有重要的意义。

5.2 人工智能实践的社会价值

从医疗保健到金融服务再到交通管理,人工智能技术正在改变着我们的生活和工作方式,我们也需要思考一下这些新技术所带来的社会影响。人工智能的普及将对就业市场产生深远的影响,虽然一些人担心自动化会导致大量失业,但是事实上,许多新的职业将会出现以适应这种趋势^[4],如数据分析师、机器学习工程师、自然语言处理专家等。人们应该积极地接受并学习新技能以应对未来的变化。

5.3 人工智能实践的生态价值

在实际应用中,人工智能不仅可以提高生产效率和降低成本,还可以为社会带来更多的福利和社会效益,从生态系统的角度来看,人工智能也具有生态价值^[5]。人工智能能够帮助我们更好地理解自然环境,并预测未来的变化趋势,它也能够促进资源利用的高效化与可持续性,减少对

生态环境的压力,通过智能化的管理手段,人工智能还能够实现更加精细化的治理模式,从而保护生态环境的完整性和稳定度,人工智能作为一种新兴的技术,有着非常重要的生态价值,在未来的发展中,应该注重其生态价值的发挥和发展,以推动人类社会的可持续发展。

6 结语

我们探讨了如何利用大数据时代的视角来推动人工智能的实践,通过对现有技术和应用的研究以及对未来趋势的预测,我们可以看到人工智能将在各个领域发挥越来越大的作用,实现这些目标需要更多的努力和创新,我们提出了一些建议,以帮助人们更好地理解并应对未来的挑战。首先,建立一个开放的数据平台,以便数据可以被共享和使用;其次,应该加强与人类之间的交互能力,使机器能够更准确地了解用户的需求;最后,我们还必须重视隐私保护,确保我们的行为不会侵犯他人的权利。在未来的发展过程中,我们将面临许多机遇和挑战,但我们相信只要我们保持积极的态度和不断探索的精神,我们就能取得更好的成果。

参考文献

- [1] 王佳林,赵心蕊,高涵.大数据时代人工智能在计算机网络技术中的应用研究[J].中国新通信,2020(2):90.
- [2] 原康.人工智能在计算机技术方面的应用探究[J].产业科技创新,2019(30):59-60.
- [3] 张嗣宏,左罗.基于人工智能的网络智能化发展探讨[J].中兴通讯技术,2019,25(2):61-66.
- [4] 孙华山,张茂兴.大数据背景下关于网络信息系统安全形势的研究[J].信息系统工程,2019,32(12):60.
- [5] 范砚博,徐振国,孙昭强.大数据时代计算机网络安全防范策略[J].中国新通信,2019,21(24):157.

Research on Real-time Security Monitoring and Auditing Technology of Big Data in Cloud Computing Environment

Miao Kong

Anhui Medical University, Hefei, Anhui, 230032, China

Abstract

With the increasing application of cloud computing and big data in various industries, which also puts forward higher challenges to data security and compliance, the research on real-time security monitoring and auditing technology of big data in cloud computing environment is particularly important. Firstly, the real-time security monitoring of big data in cloud computing environment is explored to reinforce the stable operation and security protection of the business system from the dimensions of system monitoring, security monitoring, business monitoring and resource monitoring. Then it reviews the aspects of data integrity audit, access control audit, log audit, and compliance audit to ensure the integrity and compliance of big data, with a view to providing a strong technical support for the future management and application of big data.

Keywords

cloud computing; big data; real-time security monitoring; auditing techniques

云计算环境中大数据的实时安全监控与审计技术研究

孔淼

安徽医科大学, 中国·安徽合肥 230032

摘要

随着云计算和大数据在各行各业的应用日益广泛,这也对数据的安全性和合规性提出了更高的挑战,因此,对云计算环境中大数据的实时安全监控与审计技术的研究显得尤为重要。探讨了云计算环境下大数据的实时安全监控,从系统监控、安全监控、业务监控、资源监控等维度,对业务系统的稳定运行和安全防护进行加固。针对数据完整性审计、访问控制审计、日志审计、合规性审计等方面进行评述,以确保大数据的完整性和合规性,以期为未来大数据的管理和应用提供了有力的技术支持。

关键词

云计算; 大数据; 实时安全监控; 审计技术

1 引言

随着信息技术的飞速发展,云计算和大数据技术已成为推动现代社会数字化进程的重要力量,云计算以其高效的资源利用、灵活的扩展性和便捷的服务模式,为企业事业单位提供了前所未有的数据处理能力。同时,大数据技术的应用也使得海量数据的收集、存储、分析和挖掘成为可能,极大地促进了科学研究、商业决策和社会治理的智能化^[1]。由于云计算环境的复杂性和大数据的多样性,数据泄露、非法访问、恶意攻击等安全威胁层出不穷,存在给相关机构带来经济和声誉受损的风险。因此,探讨云计算环境中大数据的实时安全监控与审计技术研究具有重要意义。

2 云计算环境中大数据的实时安全监控研究

2.1 系统监控

在云计算环境中,大数据的实时安全监控确保了大数据服务能够稳定、可靠地运行,而系统监控作为实时安全监控的核心部分,涵盖了设备状态与运行、实时性,以及可视化与智能化等多个方面。系统监控通过持续收集和分析设备的性能指标、健康状态等信息,能够迅速发现设备故障或潜在的性能瓶颈,确保云计算服务的稳定性和高可用性,一旦发现问题,系统监控能够立即触发响应机制,如发送警报通知、执行自动修复等,从而迅速解决问题,减少服务中断时间^[2]。系统监控采用实时数据分析技术,能够即时获取设备的状态信息和运行数据,并在短时间内完成大量数据的处理和分析,这使得系统监控能够迅速发现异常情况并做出响应。此外,系统监控通过可视化界面,管理员可以直观地查看设备的状态信息、性能指标等,从而更加方便地进行管理和维护。同时,基于大数据分析和机器学习技术,系统监控

【作者简介】孔淼(1981-),男,中国安徽合肥人,本科,工程师,从事大数据、信息安全研究。

能够对设备的运行数据进行深度分析,发现潜在的问题和趋势,并给出相应的建议和改进措施,这种智能化分析不仅提高了管理的效率,也进一步增强了云计算环境的安全性和稳定性。通过系统监控的实时性、可视化与智能化特性,可以确保大数据服务的稳定性和高可用性,从而为用户提供更加优质、可靠的服务体验。

2.2 安全监控

随着数据量的不断增长和云计算服务的广泛应用,安全威胁也日益多样化和复杂化,因此,实施有效的安全监控措施对于保护大数据的安全至关重要。

安全监控的核心在于实时分析系统日志、网络流量等关键数据,以便检测和评估系统运行过程中遭遇的各类安全问题,包括但不限于网络攻击、恶意软件入侵、数据泄露等。安全监控通过实时分析,能够迅速捕捉到异常行为或潜在威胁的迹象,从而确保在威胁发生之前或初期阶段就能够得到及时的处理^[3]。在安全监控过程中,安全监控技术运用先进的算法和模型,对数据进行深度分析和挖掘,以发现潜在的安全威胁,一旦检测到异常情况,系统能够立即触发相应的应对措施,这些应对措施可以是自动的,如隔离受感染的设备、阻断恶意流量、关闭部分网络端口等,也可以是手动的,如启动应急响应机制、通知管理员进行人工干预等。此外,实时分析与应对的能力使得安全监控系统能够迅速响应安全威胁,减少潜在损失。

2.3 业务监控

业务监控通过实时监控云平台上的业务状态,能够实时获取业务运行的关键指标,如响应时间、吞吐量、错误率等,这些指标直接反映了业务的运行状况,是评估业务性能的重要依据,一旦业务监控系统发现业务运行出现异常,如响应时间延长、错误率升高等,它能够立即发出警报,并触发相应的处理机制。另外,业务监控通过对业务运行数据的深入分析,业务监控系统能够发现业务性能的瓶颈和潜在问题。系统可以提供相应的优化建议,如调整网络配置、优化数据库查询语句、改进代码性能等^[4]。业务监控的另一个重要价值在于提升用户体验。用户是业务服务的最终接受者,对于业务的性能和稳定性有着直接的感受。通过业务监控,业务部门可以及时发现并解决业务运行中的问题,减少业务中断和故障发生的概率,从而提升用户体验。

2.4 资源监控

资源监控的主要任务是对云计算环境中各种资源,如CPU、内存、存储等使用情况和可用性进行实时监控。通过收集和分析这些资源的使用数据,资源监控系统能够提供详细的资源使用报告和趋势分析,帮助信息管理部门了解资源的实际使用情况,以及是否存在资源浪费或资源瓶颈等问题。在实时监控的基础上,资源监控系统还能够及时发现资源不足以及资源被滥用等问题,当系统检测到某个资源的使用量接近或超过其容量限制时,它会立即发出警报,提醒管

理员采取相应的措施,如增加资源容量、优化业务负载等^[5]。同时,资源监控系统还能够监控资源的滥用情况,在这种情况下,系统可以采取限制资源使用、隔离问题用户或应用程序等措施,以确保资源的公平分配和高效利用。除了及时发现和解决问题外,资源监控还能够进行资源的分配和优化。通过分析资源的使用数据和业务负载情况,资源监控系统可以预测未来的资源需求,并提前进行资源的分配和调度。这不仅可以避免资源的过度浪费和瓶颈问题,还可以提高系统的响应速度和可靠性,从而提升用户体验和业务效率。通过实时监控和管理各种资源的使用情况和可用性,资源监控系统能够确保系统的稳定运行和高效资源利用,为单位的发展提供有力的支持。

3 云计算环境中大数据的审计技术研究

3.1 数据完整性审计

在云计算环境中,大数据的审计技术研究对于确保数据的安全性和完整性至关重要。特别是在数据完整性审计方面,对数据源进行验证和跟踪数据变化历史是不可或缺的两个环节。在云计算环境中,大数据来源于多个不同的数据源,包括内部系统、外部合作伙伴以及公共数据集等,为了确保数据的真实性和可靠性,审计系统必须能够对数据源进行严格的验证,这包括确认数据源的身份、检查数据源的授权和访问控制机制,以及验证数据的完整性和一致性等,通过数据源验证,可以确保存储在云平台上的大数据是真实可靠的,为后续的数据分析和应用提供坚实的基础。在云计算环境中,大数据的存储和传输往往是动态和频繁的。因此,审计系统必须能够跟踪数据在云平台上的变化历史,确保数据的完整性和可追溯性,通过记录数据的创建、修改、删除等操作历史,审计系统可以及时发现数据异常和潜在的安全风险,如数据篡改、非法访问等。同时,数据变化跟踪还可以帮助管理员了解数据的使用情况和趋势,为数据管理和优化提供有价值的参考信息。通过数据源验证和数据变化跟踪等技术手段,可以确保存储在云平台上的大数据的真实性和完整性,为机构的数据安全提供坚实的保障。

3.2 访问控制审计

随着数据量的不断增长和云计算服务的广泛应用,如何有效地管理用户权限、监控和记录用户的访问行为,以及及时发现并报告异常行为,成为信息管理部门关注的焦点。审计系统应能够精确地定义和管理用户对大数据的访问权限,意味着系统需要建立严格的权限控制机制,确保只有经过授权的用户才能访问和操作特定的数据资源,通过细粒度的权限设置,系统可以限制用户对数据的访问范围、操作类型和时间等,从而有效防止数据泄露和滥用。审计系统应能够实时监控和记录用户的访问行为,包括记录用户的登录信息、访问时间、访问路径、操作内容等。通过记录这些信息,系统可以追溯用户的操作历史,了解数据的使用情况,

并发现任何潜在的安全风险。同时,这些数据还可以用于生成用户行为分析报告,为决策层提供更深入的数据洞察和决策支持。审计系统应能够识别并报告任何异常或可疑的访问行为,包括未经授权的访问、异常频繁的访问、非工作时间的访问等。通过及时发现这些异常行为,系统可以迅速采取应对措施,如限制用户访问、触发警报通知等,从而防止潜在的安全威胁。通过用户权限管理、实时监控和记录用户访问行为以及异常行为检测等手段,审计系统可以确保只有授权用户才能访问和操作数据,并及时发现和应对潜在的安全风险。

3.3 日志审计

一个高效的日志审计系统不仅能够收集云平台上的各种日志数据,还能进行深入的分析,从而及时发现并解决潜在的安全问题。云平台上的系统日志、应用日志、安全日志等包含了丰富的运行信息和安全事件数据,一个全面的日志审计系统需要能够实时收集这些日志数据,并进行分类、存储和备份,通过收集这些日志数据,机构可以全面了解系统的运行状况和安全状况,为后续分析和审计提供数据支持。在收集到日志数据后,审计系统需要利用先进的分析技术和算法,对日志数据进行深入的分析和挖掘,通过分析,系统可以识别出各种安全事件,如未授权访问、数据泄露、恶意攻击等。同时,系统还可以对日志数据进行关联分析和趋势预测,发现潜在的安全风险和威胁。在识别出安全事件后,日志审计系统需要及时采取相应的应对措施,包括向管理员发送警报通知、启动应急预案、限制用户访问等。通过及时地响应和处理,可以迅速控制安全事件的扩散和影响范围,降低安全风险。通过收集、分析和审计云平台上的日志数据,信息管理部门可以及时发现并解决潜在的安全问题,保障系统的稳定运行和数据的安全可靠。因此,建立一个高效、全面的日志审计系统对于云计算环境的安全管理至关重要。

3.4 合规性审计

随着数据保护法规的日益严格和隐私保护意识的提高,合规性审计成为各类机构不可或缺的一项工作。云平台上的大数据涉及到用户的个人信息、商业机密等敏感数据,因此必须确保这些数据在使用、存储和传输过程中符合相关法规

和规定。审计系统应具备强大的法规遵从性检查功能,能够自动扫描和分析大数据,检查其是否符合数据保护法、隐私法等法规的要求,一旦发现数据存在不合规的情况,审计系统应立即报告并采取相应的措施,如数据脱敏、加密存储等,以确保数据的合规性。审计报告是评估大数据合规性状况的重要依据,也是机构向相关方展示其合规性管理能力的重要手段,审计系统应能够自动生成详细的审计报告,包括审计结果、建议改进措施等,这些报告应包含数据的来源、存储位置、访问权限、使用情况等关键信息,并对数据的合规性状况进行客观评估。通过审计报告,信息管理部门可以清晰地了解自身在大数据合规性管理方面的优势和不足,并据此制定改进计划和措施。通过法规遵从性检查和审计报告生成等功能,审计系统可以帮助信息管理部门确保大数据的合规性,降低合规风险,并提升单位在数据保护方面的形象和信誉。

4 结论

论文深入研究了云计算环境中大数据的实时安全监控与审计技术,有效应对了数据安全和合规性方面的挑战。通过系统、安全、业务和资源的全面监控,确保了大数据的稳定运行和严密防护。同时,数据完整性、访问控制、日志审计和合规性审计等技术的运用,进一步保障了大数据的完整性和合规性。这些研究不仅可以显著提升大数据在云计算环境中的安全性和可靠性,还为大数据的管理和应用提供坚实的技术基础。展望未来,我们将继续深化研究,为大数据的安全与合规保驾护航。

参考文献

- [1] 叶剑虹.大数据分析的制造系统实时安全监控机制平台[D].泉州:华侨大学,2022.
- [2] 陶宏桥.大数据时代下审计风险的影响因素及防范措施研究[J].中国乡镇企业会计,2024(3):166-168.
- [3] 欧雨晴.从传统审计到大数据审计:转型策略与路径探讨[J].现代审计与会计,2024(2):4-6.
- [4] 过艳雯.大数据、云计算技术对审计的影响[J].中国市场,2023(9):187-189.
- [5] 丁晓蓉.大数据审计环境中的审计技术与方法探析[J].财会学习,2023(4):97-99.

Research on Plagiarism Detection Methods for Open-ended Programming Questions in Higher Education Self-study Exams

Rui Li Yixiong Zhang Qiang Guo Yu Wang Yan Bai

School of Continuing Education, Xi'an Jiaotong University, Xi'an, Shaanxi, 710049, China

Abstract

The software program design course practice assessment of higher education self-study examination usually adopts open programming questions, and plagiarism detection is an important task to evaluate students to submit programming homework. This paper proposes a deep learning-based similarity detection model to solve the problem of cross-morphological similarity detection between source code and binary code. The control flow graph information is extracted through Joern and IDA Pro tools, and multimodal features such as sequence token, character font and control flow graph structure are fused, which are processed by a semantic encoding network to generate representative vectors. This paper introduces a contrast-learning model for extracting semantic features of binary code. Triplet loss-guided representation learning is used to adjust the feature representation of the source code with the binary code to achieve code similarity detection across morphologies. Experimental results show that the proposed method maintains a high accuracy of 93.21% in the presence of compilation interference. Future research directions include further analysis of the contributions of each mode feature and exploring the possibility of introducing new modes for similarity detection.

Keywords

practical assessment; programming problems; plagiarism detection; pre-trained mode

高等教育自学考试实践考核中开放式编程题的抄袭检测方法研究

李睿 张轶雄 郭强 王昱 白雁

西安交通大学继续教育学院, 中国·陕西 西安 710049

摘要

高等教育自学考试的软件程序设计课程实践考核通常采用开放式编程题目, 抄袭检测是评估学生提交编程作业的重要任务。论文提出了一种基于深度学习的相似性检测模型, 以解决源代码与二进制代码之间的跨形态相似性检测问题。通过Joern和IDA Pro工具提取控制流图信息, 并融合序列token、字符字面量和控制流图结构等多模态特征, 这些特征经过语义编码网络处理生成代表向量。论文介绍了一个对抗编译干扰的对比学习预训练模型, 用于提取二进制代码的语义特征。采用三联体损失引导的表示学习, 调整源代码与二进制代码的特征表示, 实现跨形态的代码相似性检测。实验结果显示, 该方法在编译干扰存在情况下能保持93.21%的高准确率。未来研究方向包括进一步分析各模态特征的贡献, 并探索引入新模态进行相似性检测的可能性。

关键词

实践考核; 编程题; 抄袭检测; 预训练模型

【课题项目】中国高等教育学会2022年度高等教育科学研究规划课题《支撑S2B2C模式的自考助学智慧服务平台建设》(项目编号: 22ZXKS0308); 中国高等教育学会2023年度高等教育科学研究规划课题《面向自考网络助学的群体协同学习研究与探索》(项目编号: 23ZXKS0305); 陕西高等教育教学改革研究项目《陕西省高等教育自学考试网络助学平台数字基地建设研究与实践》(项目编号: 23JG001)。

【作者简介】李睿(1983-), 男, 中国陕西汉中, 博士, 高级工程师, 从事智慧教育与数字化考试研究。

1 引言

高等教育自学考试实践考核评估学生运用专业知识解决实际问题的能力。开放式题目有效评估综合能力、创新思维和实际应用能力, 适用于实践考核, 特别是软件程序设计课程。这类题目没有固定答案, 允许多种解决方案, 更注重方法、创新和代码质量。

开放式编程题评估考生的算法、数据结构、软件工程和用户界面设计能力, 鼓励创意和独特的解决方案。它们不仅评估个人技能, 也用于团队项目和协作能力。在这类考核中, 抄袭检测是一个重要且具有挑战性的任务。开放式题目允许多种解决方案, 结构和实现上可能差异较大, 但也存在

提交他人代码的风险。

有效的抄袭检测机制维护考核公正性和完整性。学生的编程作业可能包括源码和调用的包或库的二进制代码。为了掩饰抄袭行为，学生可能将部分源码编译为二进制代码，增加检测难度。因此，需要同时考虑源代码和二进制的抄袭检测。

代码相似性分析是检测抄袭的核心技术，备受工业界和学术界关注。相似性分析揭示作业间的结构、逻辑和文本相似度，判断潜在抄袭行为。当前技术包括文本比较、语法树比较、标记序列比较和代码“指纹”比较，但在同时检测源码和二进制代码时，传统方法会失效。

针对这种跨形态代码相似性检测的研究迫在眉睫。论文提出一种方案，使用 CodeBERT 预训练模型提取源代码语义表示向量，采用对比学习的预训练语义编码网络处理二进制代码，进行多模态特征融合，解决跨形态代码相似性比较的复杂性。论文第二章介绍相关工作，第三章探讨预处理方法，第四章探讨语义编码方法，第五章探讨特征融合和相似性检测，第六章总结与展望。

2 相关研究

2.1 特定形态下的代码相似性检测方法

2.1.1 面向源代码的代码相似性检测方法

代码相似性检测技术发展迅速，已有较为成熟的项目。例如，Moss (Measure of Software Similarity) 广泛应用于教育领域，支持多种编程语言，用户提交代码后可接收相似性报告，用于预防和检测作业抄袭。

跨语言的源代码相似性检测技术也在不断成熟，例如 JPlag，专为教育领域设计，用于发现和防止学生作业抄袭。JPlag 支持 Java、C++ 和 Python 等多种语言，能够分析这些语言之间的代码相似性，提供网页界面展示匹配结果，适用于不同编程语言项目的抄袭检测需求。

2.1.2 面向二进制代码的代码相似性检测方法

二进制代码相似性研究起步较晚。Baker 及其团队在 20 世纪末提出了一种压缩和比较可执行代码差异的方法，并开发了 EXEDIFF 工具。随着领域的发展，形成了六种主要检测技术，包括基于属性度量、代码语法、代码结构、代码语义、行为分析和代码表示学习的方法。例如， α Diff 工具通过字节流信息构建特征向量并判断相似性；Asm2Vec 采用自监督学习框架，将汇编指令序列转换为嵌入向量，通过余弦距离评估相似性。

2.2 跨形态的代码相似性检测方法

二进制代码由源代码经编译器编译而成。由于编译过程中丢失了变量名和树结构信息，取而代之的是指令、编译器优化和寄存器信息，源代码和二进制代码在字面表示上差异巨大。跨形态代码相似性检测需要解决这些差异，复杂性较高，因此相关研究较为罕见。

目前的方法主要依赖于提取代码的字面特征，如字符

串、整数字面量和递归结构，通过匹配算法评估相似度。然而，这种方法有局限性：函数代码中的字面量信息有限，忽略深层语义特征会影响准确性；且特征选择和工程需要丰富的专家经验，过程耗时且效率低。2019 年，中国科学院信息工程研究所推出了 B2SFinder 技术，检测商业软件对开源项目的重用。此外，CodeCMR 使用独立的神经网络模型从源代码和二进制代码中学习表示向量，通过三元组损失网络实现跨模态融合，在函数级别实现跨形态相似性匹配。

现有方法主要使用传统字符匹配算法来比对字面量特征，存在效率低和准确度有限的问题。由于源代码和二进制代码存在本质差异，跨形态代码相似性研究面临重重困难。设计一个融合多种模式的跨形态代码相似性检测方案，实现源代码与二进制代码之间的相似性检测任务，尤为重要。

3 代码原始表示的抽取与预处理

对于代码，粗粒度信息包括整数字面量和字符串字面量，可用于表示代码。代码的序列 token、控制流图和抽象语法树信息也可表示代码。代码相似性检测的第一步是分别提取源代码和二进制的这些特征信息。

3.1 源代码原始表示抽取及预处理

论文聚焦函数级别的相似性检测。虽然函数级相似性检测应用广泛，但代码量减少使得特征变得稀缺。为了提高模型准确性，需要提取更细致的特征，并采用匹配的表示学习模型。

源代码与二进制代码之间，字符串保持一致，意味着相似代码对包含相同的字符串字面量。尽管字符串信息较少，但可作为补充信息。代码的 token 序列表示源代码与二进制代码中的序列和指令顺序，是细粒度信息。此外，代码结构特性是相似度检测的关键因素，显示代码执行时的调用流程，具有抗混淆能力。

①字符串字面量抽取：遍历源函数，遇到双引号表示发现字符串字面量，将其内容提取并单独存储。存储时，将字符串内部空格替换为“-”，字符串间用空格分隔。若源函数不含字符串信息，则返回空值。

② token 序列信息抽取：将源函数转化为 token 序列，删除与代码语义无关的内容，如注释和空行，然后分词并标准化处理，生成适合模型训练的 token 序列。

③结构信息抽取：代码结构特征蕴含丰富的代码语义信息，能洞察代码执行过程中的调用流。本研究选用控制流图代表代码结构信息，并将其融入相似性检测模型。控制流图展示程序处理过程中的所有可能执行路径。若源代码与对应二进制代码相似，则逻辑顺序和跳转流程也应相似，其控制流图信息也会相似。

为提取源代码的控制流图信息，论文利用轻量级开源工具 Joern，该工具能基于静态源代码块自动解析并生成各类代码属性图，适合函数级别的相似性分析任务。

3.2 二进制代码原始表示抽取及预处理

3.2.1 字符串字面量抽取

使用 IDA Pro 反汇编工具获取二进制代码中的字符串字面量。通过查看“Strings”窗口内的所有字符串信息，提取目标代码中的字符串字面量。

3.2.2 token 序列信息抽取

使用 IDA Pro 工具反汇编二进制代码，将反汇编后的函数代码按行为单位存储。将空格替换为“-”作汇编指令内部的分隔符。将所有整数值替换为“dig”标识符，减少具体数值对相似性分析的影响。将不同通用寄存器名称统一替换为“reg”表示。一条汇编指令视为一个 token，指令的 token 序列通过空格分隔。这一从原始汇编代码到目标代码 token 序列的转换流程，如图 1 所展示。

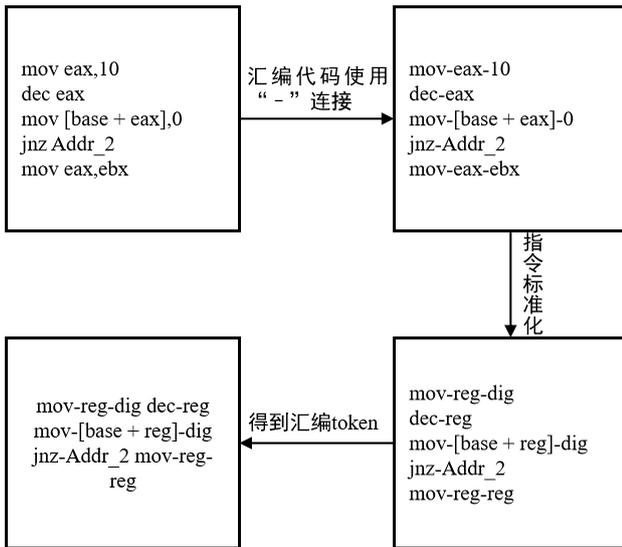


图 1 汇编 token 处理流程

3.2.3 结构信息抽取

IDA Pro 不仅反汇编二进制代码，还能提取二进制函数

的控制流图（CFG）信息。通过分析指令的跳转逻辑，函数被划分为基本块，这些基本块构成控制流图的节点，并根据跳转关系连接这些节点形成控制流图的边。通过记录每个节点的基本代码块和节点间的跳转关系，完成对二进制代码控制流图信息的抽取。

在存储控制流图（CFG）信息时，逐行记录每个节点所代表的基本代码块，并使用独立行描述节点间的跳转关系。例如，图 2 中展示的一段汇编代码及其控制流图信息，通过分析指令间的跳转逻辑，识别出函数通过分支结构实现功能。

4 代码语义编码网络

源代码与二进制代码差异大，不能直接比较相似性。语义编码网络能将文本转为反映原信息的向量，适用多种任务。代码信息有其语义和语法。本研究为源代码和二进制代码构建语义编码网络，将特征信息转化为包含语义和语法的向量。针对不同模态特征，相似性检测模型需设计特定的语义编码网络，以提取和表达这些模态特征向量。

4.1 基于预训练模型增强的源代码语义编码

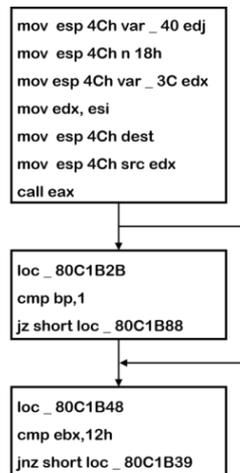
BERT (Bidirectional Encoder Representations from Transformers)，由 Google 于 2018 年提出，是 NLP 领域的一大创新。它基于 Transformer 双向编码器，通过自监督在大数据上预训练，深入学习文本的上下文关系，掌握复杂语境。BERT 具有出色的迁移学习能力，可通过微调适应具体 NLP 任务，处理复杂语义和长距离依赖问题。

Feng 及其团队首次将 BERT 引入代码处理领域，提出了 CodeBERT 模型。CodeBERT 结合了代码和相关注释的双模态数据进行训练，展现出卓越的泛化能力，适用于多语言代码检测任务，包括 Java、Python、Go 和 C 语言等。论文采用 CodeBERT 模型提取源代码序列的特征。

```

    mov esp 4Ch var_40 edj
    mov esp 4Ch n 18h
    mov esp 4Ch var_3C edx
    mov edx, esi
    mov esp 4Ch dest
    mov esp 4Ch src edx
    call eax
    loc_80C1B2B
    cmp bp,1
    jz short loc_80C1B88
    loc_80C1B48
    cmp ebx,12h
    jnz short loc_80C1B39
  
```

(a) 二进制代码



(b) 二进制代码控制流图

图 2 目标代码及其控制流图信息

CodeBERT 遵循 BERT 架构, 含 12 个编码层, 每层含 12 个自注意力单元 (尺寸 64) 和全连接层, 通过大规模语料训练优化。应用 CodeBERT 于源代码特征表征时, 以代码 token 为输入, 通过内置词嵌入模块和特征提取部分, 将代码序列转为表征性特征向量。

4.2 基于对比学习预训练的二进制代码语义编码

编译器配置差异会影响二进制代码相似性分析。论文提出一种孪生网络和对比学习的二进制语义编码模型, 减小同源代码编译结果的差异, 提升模型鲁棒性和抗干扰能力。

4.2.1 二进制样本对构建

基于代码 token 嵌入的相似性构建样本对。具体地, 假定两段二进制代码通过 token 嵌入后转为向量 ϑ_a 与 ϑ_b , 使用余弦距离衡量相似度, 公式如下:

$$Dis(Bin_a, Bin_b) = cosine(\vartheta_a, \vartheta_b) = \frac{\vartheta_a \cdot \vartheta_b}{\|\vartheta_a\| \|\vartheta_b\|} \# \quad (1)$$

$$Similarity(Bin_a, Bin_b) = 1 - Dis(Bin_a, Bin_b) \# \quad (2)$$

设置合适的相似性阈值对数据代码对构建至关重要。不相似阈值过低或过高、相似阈值过低会影响模型训练和样本对区分, 降低准确性。论文中相似性得分 > 0.8 表示相似, < 0.3 表示不相似。

确定正负样本对阈值后, 构建二进制代码对数据集, 每个样本三元组表示为 $(Bin_a, Bin_b, label)$, $label = 1$, 表示相似 (正样本), $label = 0$ 不相似 (负样本)。数据集采用

GCC、ICC、Clang 编译器的 O0、O3 优化选项编译的汇编代码, 包含 125,692 对正负样本, 其中正样本 62,352 对, 负样本 63,340 对。

4.2.2 二进制代码编码模型对比预训练

使用 GCC 编译器 O0 优化选项编译的二进制代码与源代码组成的对在 CMSim 模型中相似性检测准确, 但换成 GCC 的 O3 优化或 Clang 的 O0 优化编译的二进制时, 检测效果下降。CMSim 模型原先只训练 GCC O0 编译的数据集, 难以适应不同编译选项导致的代码差异。为减少编译选项影响, 引入基于对比学习的二进制代码语义编码网络, 通过学习同源二进制代码对的共性和非同源对的差异, 提高编码效果。

对比学习驱动的二进制语义编码网络模型采用 Text CNN、GCN 和 Bi-LSTM 分别提取 token 序列、结构和字符串字面量特征, 如图 3 所示。论文利用孪生网络简化训练, 目标是缩小同源二进制代码对距离, 扩大非同源对距离。

以源代码通过 GCC 的 O0 和 O3 优化编译得到的二进制代码 g_a 与 g_b 为例, 孪生网络处理这对代码, 通过表示学习模型提取特征向量并融合得到最终表示 x_a 和 x_b 。模型在提取三种模态特征后, 添加网络模块利用对比学习优化编码。定义相似性以保证对称性 $sim(x_a, x_b) = sim(x_b, x_a)$, 然后通过向量拼接、多层感知机处理, 并通过 sigmoid 函数得到 $[0,1]$ 区间的相似性结果。使用 Cosine loss 计算损失, 训练得到的网络可应用于跨形态代码相似性检测。

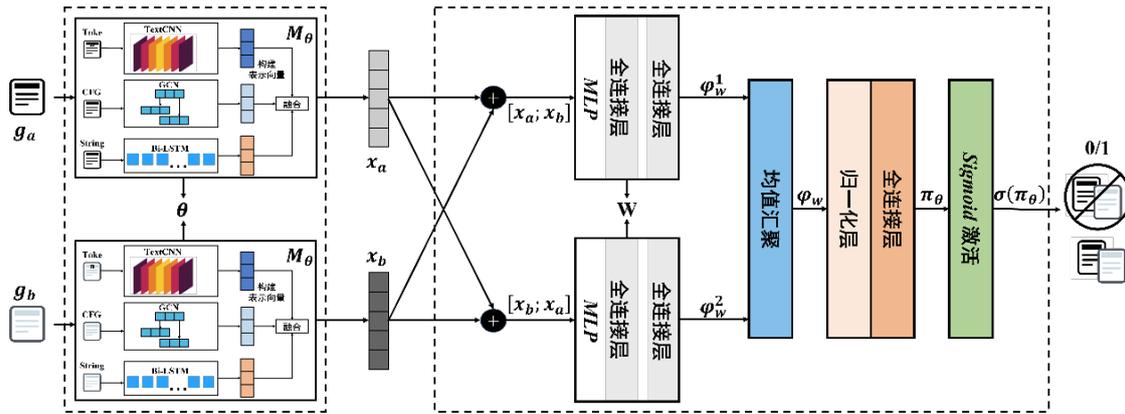


图 3 二进制代码对比学习模型

5 基于多模态特征融合的代码相似性检测

不同代码表示形式可以视为代码的单模态特征信息。多模态特征融合将各种单模态特征映射到共享语义子空间, 便于整合。多模态表示构建处理和聚合多种模态信息的模型^[12], 分为联合表示和协调表示两类。

5.1 基于联合表示策略的形态内表示向量融合

通过不同语义编码网络编码代码的字符串字面量、token 序列和控制流图特征, 得到三种模态的特征向量。联合表示将这些单模态信息映射到同一多模态空间中。实现联合表示

的方式如下:

$$x_m = f(x_1, x_2, x_3, \dots, x_n) \# \quad (3)$$

式 (3) 中, f 为融合机制, 它可以是个深度神经网络, x_1 、 x_2 等符号分别是指各模态信息经过表示学习生成的模态特征向量, x_m 是经过联合表示方法完成特征融合的融合向量。

对源代码函数 c_1 , 提取 token、字符串字面量和结构特征, 用适配的模型 α_1 、 α_2 、 α_3 训练得到各模态特征向量。

$$c_1^{tok} = \alpha_1(c_1) \# \quad (4)$$

$$c_1^{str} = \alpha_2(c_1) \# \quad (5)$$

$$c_1^{cfg} = \alpha_3(c_1) \# \quad (6)$$

其用 c_1^{tok} 、 c_1^{str} 、 c_1^{cfg} 分别表示源代码 c_1 的不同模态信息经过适配的语义编码网络生成的表示向量，使用联合表示方法融合这三个向量，其公式如下：

$$c_s = f(c_1^{tok}, c_1^{str}, c_1^{cfg}) \# \quad (7)$$

式 (7) 中， c_s 代表经过联合表示生成的源代码融合向量； f 为模态特征融合机制。

对于某一段目标代码片段 c_2 ，针对该函数代码的 token 序列特征，字符特征及结构特征信息设计适配的表示学习模型 B_1 、 B_2 和 B_3 ，经过模型的训练可以得到二进制代码各个模态的特征向量。

$$c_2^{tok} = \beta_1(c_2) \# \quad (8)$$

$$c_2^{str} = \beta_2(c_2) \# \quad (9)$$

$$c_2^{cfg} = \beta_3(c_2) \# \quad (10)$$

其用 c_2^{tok} 、 c_2^{str} 、 c_2^{cfg} 分别表示二进制代码 c_2 的不同模态信息经过适配的语义编码网络生成的表示向量，使用联合表示融合这三个向量，其公式如下：

$$c_b = f(c_2^{tok}, c_2^{str}, c_2^{cfg}) \# \quad (11)$$

其中 c_b 为经过联合表示后生成的二进制代码融合向量。

特征融合方法多样，主要包括逐点相加和直接拼接两种。论文采用向量直接拼接的方法来完成多模态代码的融合任务。定义代码模态信息的维度分别为 m 、 n 和 k ，则对于代码的三种模态特征信息可以记作： $c^{tok} \in \mathbb{R}^n$ 、 $c^{str} \in \mathbb{R}^m$ 和 $c^{cfg} \in \mathbb{R}^n$ ，使用向量拼接法实现特征融合得到的向量为 $Con_s = [c^{tok}; c^{str}; c^{cfg}]$ ， $Con_s \in \mathbb{R}^{n+m+k}$ ，使用该方法完成表示向量融合任务得到的特征向量 Con_s 的维度为 320。

定义 c_s 与 c_b 分别作为源代码及二进制代码的融合强语义代码表示向量。对于融合机制 f 的设计，考虑到不同模态的代码特征表示是从代码的各个角度对代码进行的互补性描述，对三种信息采用简单的线性融合方式即：

$$c_s = [c_1^{tok}; c_1^{str}; c_1^{cfg}] \# \quad (12)$$

$$c_b = [c_2^{tok}; c_2^{str}; c_2^{cfg}] \# \quad (13)$$

5.2 基于协调表示策略的形态间表示向量对齐

协调表示在多模态学习中，通过相似性约束分别处理单模态信息，使其映射到中间语义空间，形成统一的表示形式。由于源代码 c_s 和二进制代码 c_b 形式不同，相似性分析需要一致的表示方式。

协调表示是指多模态学习或数据融合中，通过在相似性约束下分别处理各单模态信息，引导它们映射到中间语义空间，使其具备统一的、具有相同维度的表示形式。因源函数代码 c_s 和二进制函数 c_b 代码存在不同的形式，所以在进行相似性分析时，需要采用一致的表示方式来处理这两种代码。

每种代码形态有一个特定投影函数 d_1 ，源代码的投

影函数为 d_2 ，二进制代码的投影函数为 d_3 。这两个投影函数将代码映射到一个相似性约束的协调中间语义空间中，即： $d_1(c_s) \sim d_2(c_b)$ 。

协调表示方法包括最小化余弦距离和最大化相关性等技术。论文使用余弦相似性度量法对齐不同形态间的表示向量，以实现基于协调表示的策略。

5.3 基于 Cosine 距离的跨形态代码相似性检测

通过计算两个向量间夹角的余弦值评估相似度，因此也称为余弦相似度或 Cosine 距离。其优点在于，向量间相似度仅取决于夹角，不受向量维度影响，提高了代码相似性检测模型的准确度。余弦值范围为 $[-1, 1]$ ，夹角越小，余弦值越接近 1，表示相似度越高。

对于相同子空间中的两个向量 $v_1(x_1, y_1)$ 和 $v_2(x_2, y_2)$ ，其相似度可依据公式 (14) 和 (15) 进行计算，其中 θ 表示向量 v_1 和 v_2 之间的夹角。

$$\text{dist}(v_1, v_2) = \cos\theta \quad (14)$$

$$\cos\theta = \frac{v_1 \cdot v_2}{\|v_1\| \|v_2\|} = \frac{x_1x_2 + y_1y_2}{\sqrt{x_1^2 + y_1^2} \sqrt{x_2^2 + y_2^2}} \# \quad (15)$$

为计算特征向量间的距离，将此公式扩展到两个 m 维向量，即计算 $X(x_1, x_2, x_3, \dots, x_m)$ 向量与 $Y(y_1, y_2, y_3, \dots, y_m)$ 向量之间的相似度，如公式 3.23 所示。

$$\text{dist}(X, Y) = \frac{\sum_{i=1}^m x_i y_i}{\sqrt{\sum_{i=1}^m x_i^2} \sqrt{\sum_{i=1}^m y_i^2}} \# \quad (16)$$

论文使用 $x = \{x_1, x_2, x_3\}$ 来代表融合后的源代码表示向量， $y = \{y_1, y_2, y_3\}$ 代表融合后的二进制代码表示向量，根据公式 (16) 来计算源代码 x 与目标代码 y 之间的距离。

在相似性检测任务中，需设定样本相似性阈值。检测分数超过阈值则代码相似，反之则不相似。论文默认阈值为 0.8，并评估其对模型的影响。通过计算损失和更新参数，完成跨形态代码相似性检测。

5.4 三联体损失引导的语义编码跨形态协同表示学习

在监督学习中，损失函数用于衡量预测值与真实值的差异。模型性能与损失函数的数值密切相关，性能越好的模型，其损失数值通常越低。

论文采用三联体损失^[3]的跨形态协同表示学习，通过映射空间度量样本间距离，让相似样本靠近，不同样本远离。三联体损失又被称为 Triplet loss，通过构建包含基础样本、正样本和负样本的数据三元组，确保基础样本与正样本之间的距离小于与负样本之间的距离，并使用距离边界 margin 来进行区分。Triplet loss 计算如 $Loss = \text{Max}(\text{Dis}(B, P) - \text{Dis}(B, N) + \text{margin}, 0)$ ，其结果分为三种情况：Easy Triplets（损失为 0，无需训练）、Hard Triplets（正样本距离大于负样本，损失大于 margin）、Semi-Hard Triplets（正样本距离小于负样本但未满足 margin，损失介于 0 与 margin 之间）。

论文利用含 125,692 对样本的数据集预训练二进制语义

编码网络，减少编译选项干扰。通过三联体损失优化模型，使得相似的源 - 二进制代码对的表示向量接近，而不相似的远离。

对于源代码和其对应的两个二进制代码版本 b_1 与 b_2 ，其中 b_1 与源代码 s 相似，而 b_2 与源代码 s 不相似。基于这些信息，可以构建一个三元组 $\langle s, b_1, b_2 \rangle$ ，并将相似度检测任务定义为：

$$\ell(\theta) = \sum_{\langle s, b_1, b_2 \rangle \in D} \max(0, \beta - \text{sim}(\psi_s, Y_{b_1}) + \text{sim}(\psi_s, Y_{b_2})) \quad (17)$$

其中， $\text{sim}(x, y)$ 为模型训练中使用的相似度度量，它通过采用余弦相似度 Cosine 来评估特征向量 x 与 y 之间的相似性，这些特征向量由代码表示学习模型生成； θ 表示模型训练过程中的待优化参数； D 是构建为源 - 二进制代码对的三元组形式的训练数据集； y 和 g 分别是为源代码和二进制代码各模态特征设计的表示学习模型的统称； b 是一个用以区分相似对与不相似对的非常小的常数，论文中设置为 0.06。

6 结语

论文采用 Joern 和 IDA Pro 工具从源代码和二进制代码中提取控制流图，将逻辑调用信息加入特征集，提高相似性检测准确率，克服传统方法的局限。通过语义编码网络，论

文从代码的序列 token、字符字面量和控制流图中提取特征，生成融合的代表向量，用余弦相似度评估代码对的相似性。针对编译条件变化导致的二进制代码差异，论文提出基于对比学习的二进制代码预训练模型，与 CodeBERT 结合应用于跨形态代码相似性检测，效果良好。

论文在跨形态代码相似性研究中仍有不足，未来可继续深耕以下方面：

- ①引入注意力机制，明确各模态的贡献比重；
- ②探索将整数信息作为新模态加入相似性检测；
- ③开发跨语言的源代码语义编码网络，专注代码语义，实现更有效的相似性检测。

参考文献

- [1] Andrew G, Arora R, Bilmes J, et al. Deep canonical correlation analysis[C]. International conference on machine learning. PMLR,2013(6):1247-1255.
- [2] Afham M, Dissanayake I, Dissanayake D, et al. Crosspoint: Self-supervised cross-modal contrastive learning for 3d point cloud understanding[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,2022(6):9902-9912.
- [3] Zhou J, Liu L, Wei W, et al. Network representation learning: from preprocessing, feature extraction to node embedding[J]. ACM Computing Surveys (CSUR),2022,55(2):1-35.

Research on the Application of Artificial Intelligence in Intelligent Manufacturing

Guangjian Ma

Henan Province Artificial Intelligence Research Institute, Zhengzhou, Henan, 450000, China

Abstract

With the rapid development of artificial intelligence technology, its application in the field of intelligent manufacturing is increasingly extensive. This study takes artificial intelligence technology as the means and intelligent manufacturing as the application scenario, and discusses the application of artificial intelligence in intelligent manufacturing through scientific theoretical research and practical application. Through the analysis of specific application cases, this study found that artificial intelligence technology can not only realize the automation of the production process and improve the production efficiency, but also realize the intelligence of the production process, and significantly improve the quality of products and the precision of production. In addition, artificial intelligence technology can also predict possible problems in the production process through big data analysis, and make solutions in advance to avoid losses in the production process. In short, the application of artificial intelligence technology in intelligent manufacturing is expected to promote the development of manufacturing to a higher level of intelligence.

Keywords

artificial intelligence; intelligent manufacturing; automation; production efficiency; big data analysis

人工智能在智能制造中的应用研究

马广建

河南省产学研人工智能研究院, 中国·河南 郑州 450000

摘要

随着人工智能技术的飞速发展,其在智能制造领域中的应用日益广泛。本研究以人工智能技术为手段,以智能制造为应用场景,通过科学的理论研究和实际应用探讨人工智能在智能制造中的应用。通过对具体应用案例的分析,本研究发现人工智能技术不仅可以实现生产过程的自动化,提高生产效率,还可以实现生产过程的智能化,显著提高产品的质量和生产的精度。另外,人工智能技术也能够通过大数据分析,预测生产过程中可能出现的问题,提前做出解决方案,避免生产过程中的损失。总之,人工智能技术在智能制造中的应用,有望推动制造业向更高水平的智能化方向发展。

关键词

人工智能; 智能制造; 自动化; 生产效率; 大数据分析

1 引言

人工智能作为一种充满未来感的技术走向,正高速发展并在全球范围内展开广泛应用。其中,智能制造作为制造业的新生态,正在大力推动生产流程的自动化和智能化,并为制造业走向更高水平提供可行性。近年来,人工智能在智能制造中的应用研究逐渐崭露头角,其潜力以及实际应用效果备受关注。本研究以人工智能技术为主线,以智能制造为应用背景,旨在深入探讨人工智能在智能制造中的具体应用,以期为理解和应用人工智能技术提供新的视角。通过对具体的生产过程和实际案例的调研分析,我们希望能够理解人工智能技术如何带来生产效率的提升,产品质量和生产精

度的显著改善,以及预测并解决可能的生产问题的能力。同时,我们也期待通过本研究的发现,在未来能为智能制造的发展趋势提出新的思考与启示。

2 人工智能技术的进展和发展

2.1 人工智能技术的概念和定义

人工智能(Artificial Intelligence, 简称 AI)技术已经成为当今科技发展的重要领域,其概念和定义随着时间的推移不断进化和完善^[1]。人工智能通常被定义为一种能够模拟和扩展人类智能的计算机技术和系统。它不仅涉及感知、推理、学习和处理自然语言等领域,还包括规划、预测以及机器人的自主操作等能力。人工智能技术的核心在于算法和数据的交互,通过大规模数据分析和复杂算法, AI 系统能够从经验中学习,进而在特定任务中表现出类似人类的智能行为。

【作者简介】马广建(1964-),男,中国河南郑州人,硕士,研究员,从事人工智能研究。

在广义上,人工智能可以分为弱人工智能和强人工智能两类。人工智能专注于特定任务,如图像识别、语音处理和数据挖掘,它在这些应用领域表现出高效的智能化特征,但不具备全面的认知能力。强人工智能则是能够执行所有人类智能行为的系统,具有自主决策和通用学习的能力,强人工智能目前仍处于理论探讨阶段,尚未实现实质性的突破。

具体而言,人工智能技术包括机器学习、深度学习、自然语言处理、计算机视觉等子领域。机器学习通过算法训练模型,使系统能够从数据中提取规律并进行预测决策;深度学习则利用神经网络技术,提升了图像识别、语音识别等领域的精度与效率。自然语言处理使计算机能够理解和生成人类语言,而计算机视觉则使机器具备“看”的能力,能够理解和解释视觉信息。

2.2 人工智能技术的发展历程

人工智能技术的发展历程始于20世纪50年代,当时的研究主要集中在计算机科学和认知科学的交叉领域。1956年,达特茅斯会议正式提出了“人工智能”这一概念,标志着人工智能研究的开端。20世纪60年代至70年代,人工智能技术经历了早期的兴起与探索,主要成就包括专家系统和初级的自然语言处理。受到硬件条件和算法基础限制,进展相对缓慢。20世纪80年代,专家系统广泛应用于医疗诊断和金融等领域,推动了人工智能技术的发展。进入21世纪,随着计算能力的提升和大数据技术的发展,机器学习和深度学习算法取得重大突破,涌现了大量应用场景。近年来,强化学习、生成对抗网络等新兴技术进一步推动了人工智能的发展,使其在自动驾驶、图像识别和智能制造等领域展现出巨大潜力^[2]。人工智能技术的发展不仅依赖于理论研究的创新,还受到计算硬件性能提升和数据资源丰富度的驱动,逐步实现了从理论探索到实际应用的跨越。

2.3 人工智能技术的主要特征和功能

人工智能技术的主要特征包括自学习、自适应能力和高精度的数据处理能力。其功能涵盖图像识别、自然语言处理、预测分析和智能决策。这些特征和功能使得人工智能能够在复杂和动态的生产环境中高效运行,提升制造业的自动化和智能化水平。

3 智能制造中的人工智能应用

3.1 人工智能在生产自动化中的应用

人工智能(AI)在生产自动化中的应用显著改变了传统制造业的运作模式。通过引入人工智能技术,生产过程得以实现高度的自动化,从而提高了生产效率、降低了人工成本及错误率。机器学习和深度学习算法被广泛应用于工业机器人、自动化装置和制造设备中,使其能够自主进行操作和决策。例如,在生产线上,AI机器人可以通过图像识别技术对产品进行质量检测,以高速、精确地筛选出不合格的产品,从而保证生产质量的稳定性和一致性。

神经网络算法在生产自动化中的功效不可忽视。该技

术通过模拟人脑的思维方式,实现对复杂生产流程的智能控制。自适应控制系统可根据实时反馈进行调整,优化生产参数,提高生产线的灵活性和响应速度。这种动态调整机制使得制造过程更加智能化,能够应对多变的市场需求和复杂的生产环境。

人工智能技术在生产自动化中的应用不仅限于单一环节,而是贯穿整个生产链条。通过工业物联网(IIoT)平台,生产设备能够实现互联互通,形成一个智能网络。数据采集和传输技术结合AI分析工具,能实时监控设备运行状态,预测潜在故障,自动排除生产障碍,确保生产系统的连续性和稳定性。

整体来看,人工智能技术推动了生产自动化的革新,实现了从传统人工操作到高度智能化操作的转变。在未来,随着技术的进一步发展和完善,AI在生产自动化中的应用将更加广泛和深入,从而推动制造业迈向更高水平的智能化和自动化。

3.2 人工智能在提高生产效率与产品质量中的应用

人工智能在提高生产效率与产品质量的应用已经显现出显著的成果。通过机器学习和深度学习,生产系统能够准确预测和适应各种生产环境,优化生产流程,实现设备的高效调度和资源的合理分配。智能机器人和机器人系统通过AI驱动的控制算法,可以独立完成复杂的生产任务,大幅减少人为操作带来的误差和时间损耗,从而提高生产效率。

在产品质量控制方面,计算机视觉技术被广泛应用于质量检测环节。通过高速摄像头和图像处理算法,可以实时监测生产过程中的每一个环节,识别缺陷并即时反馈,保证产品的一致性和高品质。人工智能还能通过大数据分析,识别产品生产过程中的潜在问题和瓶颈,进而提出改进方案,使得产品质量不断提升。

利用人工智能,智能制造系统不仅实现了生产过程的高度自动化和智能化,还通过数据驱动的方式推动了全流程的质量控制和优化,显著提升了整体运行效率和产品可靠性。

3.3 人工智能在预测和解决生产问题中的应用

人工智能技术通过大数据分析和机器学习,有效预测生产过程中可能出现的问题,帮助预先制定解决方案,避免生产过程中的损失。这些技术能够实时监测生产线的各项指标和设备运行状态,快速识别潜在的故障和瓶颈,并依据历史数据和模式进行动态调整。例如,通过预测性维护,及时发现设备故障隐患,减少停机时间,提高设备利用率和生产效率。通过优化生产调度方案,实现资源的最优配置,保障生产过程的连续性和稳定性。人工智能在这一领域的应用显著提升了问题解决的效率和准确性,推动了智能制造的持续发展。

4 人工智能推动智能制造发展的前景展望

4.1 人工智能技术在智能制造中的应用优势

人工智能技术在智能制造中的应用优势体现在多个方

面,且呈现出显著的技术创新和产业升级潜力。人工智能技术通过自动化生产系统,能够高效地处理大量复杂的生产任务,显著减少人为干预^[3]。智能制造过程中,人工智能驱动的自动化系统能够全天候、不间断地运行,提升设备利用率和生产效率,降低人工成本和操作失误的风险。这种无缝的自动化操作可以达到极高的精确度,确保生产过程的一致性和产品质量的稳定性。

在提高生产效率与产品质量方面,人工智能技术也展现出巨大的优势。机器学习和深度学习算法可以通过大数据分析,优化生产工艺流程和资源配置。借助实时数据监控和分析,生产系统能够自适应调整参数,实现最优的生产状态。这不仅提高了生产效率,还能有效减少原材料浪费,降低生产成本。

人工智能技术在预测和解决生产问题上也表现出极高的敏捷性和准确性。通过对海量历史数据的分析,人工智能系统可以预测设备故障、质量缺陷等潜在问题,并提供预防性维护建议。这种提前预警机制为企业提供了足够的时间和空间来调整生产计划,避免生产停滞和经济损失。

总的来说,人工智能技术在智能制造中展现出的应用优势,不仅体现在生产效率和产品质量的提升上,还通过其卓越的数据分析能力,提升了生产系统的智能化水平和问题预见性。这些优势为制造业的智能化转型提供了强有力的技术支持,推动了产业的持续创新与发展。

4.2 人工智能技术推动智能制造向高级智能化发展的可能性

人工智能技术在智能制造领域的应用,显著推动了该领域向高级智能化发展的进程。通过深度学习、数据挖掘和机器学习等技术,生产设备能够自我优化、自主决策,并通过实时数据分析调整生产参数,提高生产的灵活性和适应性。这种能力使得生产系统不仅可以适应小批量、多品种定制化生产的需要,满足个性化制造的趋势,还大幅度提升了生产效率和降低了生产成本。现代化生产线利用人工智能进行诊断和维护,通过预测性维护技术,设备的故障率和停机时间得以大幅减少,从而延长了设备的使用寿命并确保了生产的连续性和稳定性。

进一步,人工智能技术还能够通过数据集成和信息共享,打破生产过程中数据孤岛,实现全产业链的智能化协同。这种智能化协同不仅包括生产环节,还延伸至供应链、物

流、销售等各个环节,形成全方位智能制造生态系统。随着5G和物联网技术的发展及应用,人工智能技术将在智能制造中更为广泛深入地应用,推动制造业朝着高度智能化、网络化和柔性化方向发展。

4.3 面临的挑战与应对策略

尽管人工智能技术在智能制造中展现出巨大的潜力,但仍然面临诸多挑战。技术成本较高使得中小企业难以大规模应用;数据隐私和安全问题亟待解决;技术人员短缺限制了人工智能的采用和发展。应对这些挑战的策略包括降低技术应用成本,加强数据安全保障,推动人工智能教育和培训,培养更多专业人才。制定相关政策和标准,对人工智能应用进行规范和引导,以促进其在智能制造中的健康发展。

5 结语

在对人工智能在智能制造中的应用进行深入研究后,我们得出以下主要结论:首先,人工智能技术不仅可以实现生产过程的自动化,提高生产效率,而且还可以通过智能化的方式显著提高产品的质量和精度;其次,人工智能技术在预测和避免生产过程中可能出现的问题方面也体现出显著的优势,这是通过大数据分析实现的;最后,研究还发现,人工智能技术在制造业的应用有望推动其向更高水平的智能化发展。然而,我们不能忽视的是,目前人工智能在智能制造中的应用仍面临一些挑战。如算法改进、数据采集和分析、硬件设备的提升等问题是需要我们进一步研究和解决的。此外,人工智能的普及和应用必然会带来一系列的社会伦理和法律问题,这也是我们未来研究的重点。总的来说,人工智能在智能制造中的应用具有广阔的前景,但我们还需要进一步深化研究,完善技术,为制造业的智能化发展贡献更多力量。本研究的成果对于理解和应用人工智能技术具有重要的理论和实践意义,并为未来智能制造的发展样态提供了参考和借鉴。

参考文献

- [1] 李涇.人工智能在汽车制造业智能制造中的应用研究[J].汽车周刊,2022(1).
- [2] 闻力生.人工智能在服装智能制造中的应用[J].纺织高校基础科学学报,2020,33(2).
- [3] 赵立增.人工智能在智能制造的应用探究[J].中国科技期刊数据库 工业A,2020(11).

Analysis of Digital Twin Technology Applications in the Military Field

Xuejiao Fan Shuang Jin

Hainan Armed Police Corps, Haikou, Hainan, 570203, China

Abstract

Under the background of the rapid development of science and technology, facing the fierce international environment and the arduous tasks of domestic reform, development and stability, the field of national security has been greatly affected, and the military field has also undergone changes. Information warfare, intelligent warfare and scientific and technological warfare will become the new combat modes of future military operations. It has gradually become an effective way to improve national defense force, military capability and national competitiveness. This paper reviews the concept, development history, key characteristics and basic structure of digital twin technology, and analyzes the application of digital twin technology in military field from four aspects: optimizing command system, enhancing equipment health prediction ability, quickly dealing with network faults, and promoting battlefield calculation.

Keywords

digital twinning; military; application analysis

数字孪生技术在军事领域的应用分析

范雪娇 金双

武警海南总队, 中国·海南海口 570203

摘要

科技迅猛发展的时代大背景下, 面对风高浪急的国际环境和艰巨繁重的国内改革发展稳定任务, 国家安全领域受到重大影响, 军事领域也随之发生变革, 信息战、智能战、科技战将成为未来军事行动的全新作战方式, 云计算、物联网、数字孪生等新兴技术的发展, 逐渐成为提升国防力量、军事能力和国家竞争力的有效途径。论文梳理了数字孪生技术的概念、发展历程、关键特征、基本架构, 从优化指挥体系、增强装备健康预测能力、快捷处理网络故障、推进战场推演四个方面分析了数字孪生技术在军事领域的应用。

关键词

数字孪生; 军事; 应用分析

1 引言

数字孪生作为近年来最具颠覆性的科学技术之一, 正日益成为各行各业的热门话题, 已经初步得到应用和研究。一方面, 在欧洲国家大型军工企业中, 数字孪生技术应用于武器装备设计、制造和运维的方面, 已经取得一定实效, 既降低成本, 也加快了生产制造效率; 另一方面, 美军为研究数字孪生体开启了“数字曼哈顿计划”, 它是新一轮军事革命的核心目标, 也是美国争夺全球军事、产业和经济领导地位的重大战略计划, 它完全满足美军分布式作战周期短暂、预算低廉的刚需。可以看到, 数字孪生技术已经成为世界大国之间军事抗衡的关键棋子, 其在军事领域的应用已然成为一个重要的研究课题。

【作者简介】范雪娇(1995-), 女, 中国陕西渭南人, 本科, 助理工程师, 从事大数据分析、人工智能和通信研究。

2 数字孪生技术的概念及发展历程

2.1 数字孪生技术的概念

目前, 数字孪生技术作为一项新技术, 概念定义没有明确的规范, 各行各业也是仁者见仁, 智者见智。但是对实体空间、虚拟空间、数据双向交换以及映射关系、决策辅助这几类重要板块的理解, 各领域不约而同达成了共识。总的来说, 数字孪生(digital twin), 是指通过数字化、物理建模等方式, 利用运行数据和平台软件, 在实体和虚拟空间形成精准映射关系、实时反馈机制, 构建虚拟世界对物理世界描述、诊断、预测和决策的新体系, 能够反映对应实体的全生命周期。

2.2 数字孪生技术的发展历程

数字孪生技术起源和发展的观点众说纷纭, 总体来看, 可大致分为5个阶段:

技术萌芽期: 建模仿真技术的应用和发展为数字孪生技术奠定了坚实的基础, 在20世纪60年代的“阿波罗”计

划中，美国国家航空航天局制作了两组完全相同的空间飞行器，一组执行发射任务，另一组被命名为“孪生体”供科学研究人员观察执行任务飞行器的运行状态。准备期间，两组飞行器均同样全程参加训练，任务开始时，孪生体与实体同步执行指令和操作，有效的飞行数据可以较为精准地反映实体的飞行条件、运行情况。

理论发展期：数字孪生的基本思想最早是在2002年，由密歇根大学 Michael Grieves 教授在 PLM (Product Lifecycle Management) 中心启动会上提出的。他认为，每个产品的完整生命周期都可以通过物理实体设备的数据，在虚拟空间构建一个能够体现该物理实体各项指标变化的设备或子系统，并建立双向动态的联系通道。这个概念在当时被命名为“信息镜像模型”(Information Mirroring Model)，也被称为“PLM 的一个理想化概念”。

概念提出期：2010年，美国国家航空航天局发布报告《建模、仿真、信息技术和处理路线图》正式提出数字孪生的概念，该报告也公布了数字孪生技术在飞行器执行任务时，对当下情况的分析、飞行状态性能变化的精准预测等方面的应用成果。

应用探索期：2011年，美空军研究实验室将数字孪生技术探索性应用于航空航天飞行器的健康管理。每次飞行任务结束时，都可以根据孪生体的状态数据，分析评估飞行器是否需要保养维修，以及能否顺利完成下次任务等等。除此之外，德国西门子、法国达索、美国通用电器等一些企业公司也在积极探索。例如，波音 777 客机在开发设计期间没有建立任何图纸模型，300 万个零部件仅依靠数字仿真来测试，测试完成后直接进行量产。根据后期数据统计，此项技术帮助波音公司缩短 40% 的研发周期，减少 50% 的返工量。

推广发展期：2020年后，世界各国纷纷将数字孪生技术纳入科技战略发展范畴，特别是近些年智能制造等概念的推进，使数字孪生技术迎来了推广应用的快速发展期。目前，杭州市利用数字孪生技术实现了“智慧城市”的管理建设，不仅可以在城市中心实时获取分析城市气象、水文、路况、事故灾害等民生信息，还可以实现信号灯的智能配置，有效缓解高峰时期的交通压力，减少市民的出行通勤时间，提高救护、应急、消防等车辆的出勤效率。

3 数字孪生技术的关键特征

数字孪生技术实现了实体空间与虚拟数字模型的精准映射和反馈，具有重要的科技发展价值，总的来说，其关键特征有以下三点。

3.1 精准映射虚实交互

数字孪生技术是仿真应用技术的发展和升级，利用智能化算法，在孪生体与实体之间形成“形神兼似”的精准映射，不仅包括行为逻辑、业务流程、物理模型，还包括参数调整所致的状态变化等等，全方位、精准、精确地在数字世

界呈现出物理实体的状态和行为。同时，实体空间与虚拟空间的演进是交互、双向进行的，孪生体实时反映物理实体的状态，还可以根据实体的反馈信息和现实需要，对实体进行干预和控制，形成双向互通信息的控制渠道。

3.2 数据驱动实时同步

数字孪生技术的主体是物理实体与行为逻辑建立的驱动模型，数据是基础，也是驱动力。孪生体是数据存储的平台，负责融合处理原始数据，驱动仿真模型各个部分动态运转，反映实体的各业务流程。数据的新鲜度直接影响孪生体与实体映射关系变化的速度，越新鲜的数据，越能驱动孪生体与实体的实时同步。

3.3 自主优化辅助决策

数字孪生的映射关系是双向的，可自主辅助实体科学决策。在运行过程中，孪生体不间断地收集归纳实体数据，并基于丰富的历史经验和先进的算法模型，通过比照、判断，动态分析实体历史态势、运行结果，科学预测未来发展趋势，适时自动调整实体的运行方向，达到自主优化生产计划、降低潜在风险、实现反馈控制和实时决策的最终目的。

4 数字孪生技术的基本架构

数字孪生技术可以帮助人们更好地理解 and 预测物理世界中的境况，这一功能的实现完全依赖于相应的技术架构，总的来说，数字孪生技术的基本架构可分为 4 个主要层级。

4.1 数据采集存储层

物理实体数据的采集存储是实现数字孪生技术的第一步，选择合适的采集工具、确定高质量的数据来源及确保数据存储设备的安全，是搭建该层级的重点。一方面，构建孪生体所需的数据量精准度高、体量大，可通过物理量算所得，也可通过各类传感器、监测设备获取，来源众多，智能的数据采集设备和适当的数据采集方式是确保采集数据质量高、准确的关键；另一方面，数据存储方式的可靠性直接影响到数字孪生技术的实现效果，设计恰当的数据库架构和管理系统可以确保底层数据安全、可信。

4.2 建模仿真层

建模仿真层实现了对物理实体的数字转化，是数字孪生技术的核心层级，其搭建的关键在于对原始数据的预处理和构建合适的数字模型。一方面，实时采集的数据可能包含缺失值、异常值、重复值等问题，容易导致建模的错误结果。通过数据预处理，可以消除原始数据中的错误和不一致，使数据更具可解释性，降低模型训练难度。另一方面，模型的搭建通常要经历三个阶段：一是选择阶段，根据建模目标 and 需求，选择合适的模型类型，如数据驱动模型、混合模型等。二是构建阶段，基于选定的模型类型，利用相关算法和工具进行模型构建，实现数字孪生模型的初步搭建。三是优化阶段，通过参数调整、算法改进等方法对模型进行优化，提高模型的精度和效率。

4.3 可视化交互层

对用户友好且易于操作的交互界面是数字孪生实现理解人类,辅助人类决策的重中之重,它是用户与机器的认知核心“桥梁”,不仅能够使机器充分了解用户需求,完成“对话”,还可以从多角度、全方位展示建模仿真成果,及时映射出辅助决策结果,清晰地展现出孪生体对应的状态、境况。

4.4 分析优化层

分析优化层是数字孪生技术的最后一个层级,通常涉及多种数据分析方法和算法,并需要结合实际情况进行综合评估。该层级主要对采集到的数据、建模结果和仿真结果进行分析,并提出相应的优化方案,需要合适的分析工具和算法。

5 数字孪生在军事领域的应用

当前,数字孪生技术正处于提质增效期,将会逐步应用于军事领域的各个方面,影响深远。总的来看,其应用主要有以下几个方面。

5.1 优化指挥体系

数字孪生技术可以利用自主学习,同步复制与实战环境一致的虚拟场景,搭建融合人员力量、武器装备、战场设施、战场环境等要素为一体的智能平台,全面优化指挥员在军事行动中的指挥信息体系。一方面,增强了现场各类信息、体系的融合能力。在数字孪生技术搭建的智能平台上,相比以往数字化、信息化的体系融合,最大的特点在于虚实映射、实时联动。各个任务分队、武器装备、物资储存,甚至是各类系统平台中的要素都可以做到即时感知、即时更新,形成“现场态势孪生体”,有效增强不同体系间的融合能力与自适应效能。另一方面,增强了指挥员现场感知能力和指挥沟通效率。数字孪生技术可以为指挥员与指挥员之间,指挥员与信息系统之间提供畅通、清晰、不间断的交流渠道,从当前技术发展的角度来说,在理想状态下这种交流包括但不限于语音、视频和触觉,可以让指挥员迅速、全面地掌握现场态势,犹如身临其境一般。同时,数字孪生技术平台还可以增强不同层次指挥对象,平行和交叉的纵横联通能力。相比于传统的军事文书的流转、处置命令下达的方式,平台可以使最高指挥员的决心意图,在紧急情况下第一时间到达各级指战员,还可以加强平行分队指挥员情报资源共享。

5.2 增强装备健康预测能力

强有力的装备支撑,是军事目标达成的重要环节。数字孪生技术通过利用大数据、云计算和神经网络等智能算法,增强装备健康的预测能力。在孪生体与实体同步完成任务操作后,可以依据孪生体的状态,分析出实体装备使用维

修的可靠数据,归纳出装备健康与储存使用地域天气、温湿度等要素之间的规律联系,建立装备健康状态、维系频率、需求关系模型。该模型可及时发现装备故障的隐患苗头,对临近发生故障或状态异常的装备预警提示,不仅提高装备运行效率,还大大缩短维修周期,减少军事装备维护巨大的资金支出。

5.3 快捷处理网络故障

以计算机为核心的信息网络已经成为军队的神经中枢,一旦遭到破坏,后果将不堪设想,数字孪生技术可以协助操作人员快速处理通信网络故障。在虚拟空间复制物理实体网络,建立与交换机、路由器等物理设备、网络架构完全一致的通信网络数字孪生体,同步开展网络传输任务时,可以通过孪生体实时观察物理实体网络的运行状态和信号流转质量,及时预测网络拥堵、信号误码等情况。当网络通信出现故障时,可以迅速定位故障点,免去人工排查,同时自适应结构组网,生成最佳抢修方案,协助操作人员快速处理。

5.4 推进战场推演

数字孪生技术可以根据任务行动和训练活动不同的任务背景,构建出真实的推演环境。采用体系建模的方法对实体类、结构类、操作类作战资源实体进行孪生建模,通过仿真推演,核算出最佳单位编组和行动策略。特别是在应对突发事件时,面对情况情报不明朗、事件爆发突然迅速的情形,战场推演就显得尤为重要,构建好的数字孪生体即可提供顺畅推演进程,圆满实现远程高效指挥。

6 结语

高质量发展是新时代的硬道理,需要新的生产力理论来指导。在信息技术爆炸发展的时代,数字孪生技术作为有力新质生产力之一,已经在实践中形成并展示出对高质量发展的强劲推动力、支撑力,用以指导其在军事行动领域的发展实践,具有强大发展动能。在未来智能化战争中,整合科技创新资源,依托数字孪生技术开展作战指挥、信息通信、战场建设和装备管理将会是大势所趋。

参考文献

- [1] 于知匀,方政.数字孪生技术及其军事应用探析[J].战略前沿技术,2024(1):14.
- [2] 庞雪凡,王振宇.数字孪生技术在智能化战争中应用[J].军事文摘,2022,4(8):23-26.
- [3] 韩将星.6G时代数字孪生在无线电监测站的应用研究[J].通信技术,2021,54(2):352-362.
- [4] 邓焯,奉祁林,赵健.数字孪生战场建设探讨[J].防护工程,2020,42(3):58-64.